

Result No.	Score	Query	Match	Length	DB ID	Description
c 1	261	GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.	CC567140	CH240_441		
		OM nucleic - nucleic search, using sw model	BY276266	BY276266		
Run on:	July 18, 2004, 14:36:45 ; Search time 1815 Seconds (without alignments) 42941231 Million cell updates/sec		BB638573	BB638573		
Title:	US-10-051-769-2		BY734796	BY734796		
Perfect score:	261		BU706546	UI-M-FOO-		
Sequence:	1 gatcaaggatggaggatccgagg.....caccctggccatcgacgttggaa 261		BQ960960	AGENCOURT		
Scoring table:	OLIGO_NUC		AK041847	Mus muscu		
	Gapext 60.0		AW013379	SP042ks W		
Searched:	27513289 seqs, 14931090276 residues		BB866050	BB866050		
Word size :	0		BX369637	BX369637		
Total number of hits satisfying chosen parameters:	55026578		BB655451	BB655451		
Minimum DB seq length: 0			BU449181	603767430		
Maximum DB seq length: 2000000000			BU463741	63366886		
Post-processing: Listing first 45 summaries			AY399655	Mus muscu		
Database :	EST:*		AK085351	AK085351		
	1: em_estba:*		AK082438	Mus muscu		
	2: em_estchum:*		BU282131	603865413		
	3: em_estin:*		AL921366	yeast2b02_x		
	4: em_estmu:*		BB660976	BH660976		
	5: em_estov:*		AK042994	Mus muscu		
	6: em_estpl:*		AY399655	Mus muscu		
	7: em_estro:*		AK085351	Mus muscu		
	8: em_htc:*		AK082438	Mus muscu		
	9: gb_est1:*		BU282131	603865413		
	10: gb_est2:*		AL921366	yeast2b02_x		
	11: gb_htc:*		BB660976	BH660976		
	12: gb_est3:*		AK042994	Mus muscu		
	13: gb_est4:*		AY399655	Mus muscu		
	14: gb_est5:*		AK085351	Mus muscu		
	15: gb_estrun:*		BU282131	603865413		
	16: em_estom:*		AL921366	yeast2b02_x		
	17: em_gss_hum:*		BB660976	BH660976		
	18: em_gss_inv:*		AK042994	Mus muscu		
	19: em_gss_pln:*		AY399655	Mus muscu		
	20: em_gss_vrt:*		AK085351	Mus muscu		
	21: em_gss_fam:*		BU282131	603865413		
	22: em_gss_mam:*		AL921366	yeast2b02_x		
	23: em_gss_mus:*		BB660976	BH660976		
	24: em_gss_dro:*		AK042994	Mus muscu		
	25: em_gss_rid:*		AY399655	Mus muscu		
	26: em_gss_phg:*		AK085351	Mus muscu		
	27: em_gss_trt:*		BU282131	603865413		
	28: gb_gss1:*		AL921366	yeast2b02_x		
	29: gb_gss2:*		BB660976	BH660976		
SUMMARIES						
RESULT 1	BF330450/c					
REFERENCE	Dias Neto, B., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P., Goldman, G.H., Carvalho A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.					
AUTHORS	1 (bases 1 to 583)					
DEFINITION	Locus: BF330450					
ACCESSION	MR2-BN0364-280800-005-f07					
VERSION	BF330450.1					
KEYWORDS	EST.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
Mammalia; Eutheria; Primates; Chordata; Craniata; Vertebrata; Euteleostomi; Catarrhini; Hominidae; Homo.						
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)					
PUBLMED	10737800					
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, Sao Paulo-SP,					

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
c 1	261	GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.	BF330450	MR2-BN0364	583 bp	mRNA linear Homo sapiens cDNA, mRNA sequence.
2	261	OM nucleic - nucleic search, using sw model	BC031019	BX345030	3069	Homo sapi
3	248	Run on: July 18, 2004, 14:36:45 ; Search time 1815 Seconds (without alignments) 42941231 Million cell updates/sec	989	BX345032	13	BX345032
4	164		944	BX345030	13	BX345030

%

Brazil			
Tel: +55-11-2704922			
Fax: +55-11-2707001			
Email: asimpson@ludwig.org.br			
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/GetHtml2.pl?i1=MR2&t2=BN0364-1)			
Seq primer: puc 18 Forward			
High quality sequence start: 10			
High quality sequence stop: 581.			
Location/Qualifiers			
source			
1. .583			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/dev_stage="Adult"			
/clone_id="BN0364"			
/note="Organ: breast normal; Vector: puc18; Site 1: Smal; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the PUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."			
FEATURES			
source			
Query Match Score 100.0%; Score 261; DB 10; Length 583;			
Best Local Similarity 100.0%; Pred. No. 4.7e-106;			
Matches 261; Conservative 0; Missmatches 0; Indels 0; Gaps 0; Oligo			
1. /			
1 GATCAAGGTGGAGCTTGGAGGACTTGGCTGCGACCAAGACGGCGCCGCTGGAGGG 60			
482 GATTAAGTGGAGTTGGAGGCTTGCGCTGCGACCAAGACGGCGCCGCTGGAGGG 423			
61 GCTGAGCCCTCGCGACGCTGTTCTGGGGAGAGGGGCTCATTAAGACCATCGGGCT 120			
422 CCTGAGCTCTGGCGAACGCTGTTCTGGGGAGAGGGGCTCATTAAGACCATCGGGCT 363			
121 CGTGCGGCAAGTGTGCCCTGCACCGGGAGGCCGATGCCCTGAAGGGGACGGCT 180			
362 CGTGCGGCAAGTGTGCCCTGCACCGGGAGGCCGATGCCCTGAAGGGGACGGCT 303			
181 GCGGGCCGCTGCCTGGAGGACTTGGCTGCGAGGGAGGTGAGTACAACGGGGCTT 240			
302 GCGCCGCCCTGCCTGGAGGACTTGGCTTGCAGGGGAGTACAACGGGGCTT 243			
241 CCACTGGCCCATGCGCTGGA 261			
242 CCACCTGGCCCATGACGGCTGGA 222			
ORIGIN			
SULT 2			
033019			
USCIS			
DEFINITION			
Homo sapiens hypothetical protein LOC118987, mRNA			
IMAGE:48217521, containing frame-shift errors.			
SESSION			
BC033019.1 GI:21542573			
HTC.			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
Klausner, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Schueler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaeffer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.J., Hsieh, F.P., Diatchenko, L., Matsusima, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustini, M.B., Toshiyuki, S., Carninci, P., Prange, C., Rahas, S.S., Loquellano, N.A., Petters, G.J.,			
AUTHORS			
REFERENCE			
1 (bases 1 to 3039)			
BC033019			
BC033019.1			
TERMS			
KEYWORDS			
SOURCE			
ORGANISM			

Abramson, R.D.,	Mullahy, S.J.,	Bosak, S.A.,	McEwan, P.J.,	
McKernan, K.J.,	Malek, J.A.,	Gunaratne, P.H.,	Richards, S.,	
Worley, K.C.,	Hale, S.,	Garcia, A.M.,	Gay, L.J.,	Hulik, S.W.,
Villalon, D.K.,	Muzny, D.M.,	Sodergren, E.J.,	Lu, X.,	Gibbs, R.A.,
Fahey, J.,	Helton, E.,	Kettman, M.,	Madan, A.,	Rodrigues, S.,
Sanchez, A.,	Whiting, M.,	Madan, A.,	Young, A.C.,	Shevchenko, Y.,
Bouffard, G.G.,	Blakesley, R.W.,	Touchman, J.W.,	Green, E.D.,	
Dickson, M.C.,	Rodriguez, A.C.,	Grimwood, J.,	Schmitz, J.J.,	Myers, R.M.,
Butterfield, Y.S.,	Krzewinski, M.I.,	Skalska, U.,	Smailus, D.E.,	
Schnerch, A.,	Schein, J.E.,	Jones, S.J.	and Marra, M.A.	
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26),	16899-16903	(2002)	
ISSN	22388257			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 3069)			
AUTHORS	Strausberg, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (21-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>			
COMMENT	Contact: MGC help desk Email: cgabps-r@mail.nih.gov			
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (Riken) CDNA Library Arrayed by: I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Institute for Systems Biology contact: <a href="http://www.systemsbiology.org">http://www.systemsbiology.org</a>			
	Anup Madan, Jessica Fahey, Erin Helton, Mark Kerteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting			
FEATURES	Location/Qualifiers			
source	1..3069			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clones="IMAGE:4841752"			
	/tissue_type="Testis"			
	/clone_id="NIH_MGC_97"			
	/lab_host="DH10B"			
	/note="Vector: pBluecript"			
ORIGIN				
Query Match	Score 261; DB 11; Length 3069;			
Best Local Similarity	100.0%			
Matches 261; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1 GATCAAGTGGAGTTCGAGGGACTGCTCAGACCAAGAGCCGCGGCGCTGCTGAGGG 60			
Db	610 GATCAAGTGGAGTTCGAGGGACTGCTCAGACCAAGAGCCGCGGCGCTGCTGAGGG 669			
Qy	611 GCTGAGCTGGAGCTGGTTCCTGGGGAGCGTGGCTCATAGACCATTGGCT 120			
Db	670 GCTGAGCTGGAGCTGGTTCCTGGGGAGCGTGGCTCATAGACCATTGGCT 729			
Qy	121 GTCGCGCCGCGCCGCCGAGGAAGCTGGCTTCAGGGGAGGTGAGTAACACGGGGCT 180			
Db	730 CGTGCGAGCAGTGCTGGCTCGGCCAGGGAGCCATGGCCCTGAGGGGGCT 789			
Qy	181 GCCCGCGCCGCCGCCGCCGCCAGGAAGCTGGCTTCAGGGGAGGTGAGTAACACGGGGCT 240			
Db	790 GCGCGCGCCGCCGCCGCCGCCAGGAAGCTGGCTTCAGGGGAGGTGAGTAACACGGGGCT 849			



REFERENCE	1 (bases 1 to 738)								
AUTHORS	Batalov,S., Beisel,K.W., Blake,J.A., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,B., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Garibaldi,M., Giassi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kurochkin,I.V., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahilstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagi-sawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Kojino,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Inotani,K., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasuno,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.								
TITLE	Unpublished Sequences from Library CHORI-240, PLATES 399 to 478 Other GSSE: CH240_441017.T7								
JOURNAL	Sequencing								
COMMENT	The British Columbia Cancer Agency Genome Science Centre 600 W 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6 Tel: 604-877-6085 Fax: 604-877-6276 Email: rholte@bcgsc.ca								
Clones are derived from the bovine BAC library CHORI-240 ( <a href="http://www.chori.org/bacpac/bovine240.htm">http://www.chori.org/bacpac/bovine240.htm</a> ). For BAC library availability, please contact Pieter de Jong ( <a href="mailto:pdejong@mail.cho.org">pdejong@mail.cho.org</a> ). Clones may be purchased from BACPAc Resources ( <a href="http://www.chori.org/bacpac/ordering">http://www.chori.org/bacpac/ordering</a> ). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Sciences Centre, Canada. Plate: 441 row: O column: 17 Seq primer: SP6 Class: BAC ends									
FEATURES	Location/Qualifiers								
SOURCE	1. 738 /organism="Bos taurus" /mol_type="genomic DNA" /strain="breed: Hereford" /db_xref="taxon:9937" /clone="CH240_441017"								
	/cell_type="Blood" /clone_lib="CHORI-240" /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI; Hereford bull Li Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"								
	/sex="Male"								
QY	Query Match Best Local Similarity 26.4%; Score 69; DB 29; Length 738; Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;								
Db	193 CCCCCGAGCTGGCCTTGAGGCCGAGGTACAAAGGGGCTTCACCTGGCCAT 252 538 CCCCGAGCTGGCCTTGAGGCCGAGGTACACGGGGCCTTCACCTGGCCAT 597								
QY	253 CGACGTGCA 261 598 CGACGTGCA 606								
Db									
RESULT 6 DEFINITION BY276266 BY276266 RIKEN full-length enriched, visual cortex mRNA linear EST 11-DEC-2002									
ACCESSION BY276266	CDNA clone K430318E15.5', mRNA sequence								
VERSION BY276266.1	GI:26466503 EST.								
KEYWORDS SOURCE	Mus musculus (house mouse)								
ORGANISM	Bukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Murinae; Mus.								
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Okada,I., Osato,N., Saito,H., Suzuki,I., Yagi,K., Tomaru,Y., Yamamoto,I., Kiyosawa,H., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Quackenbush,J., Schnitziel,L.M., Matsuda,H., Hume,D.A.,									
LOCUS_DEFINITION	By276266 RIKEN full-length enriched, visual cortex								
VERSION	Mus musculus								
KEYWORDS									
ORGANISM									
REFERENCE 1 (bases 1 to 437)									
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,									

```

/dev stage="3" days neonate"
/lab_host="DH10B"
/clone.lib="RIKEN full-length enriched, 3 days neonate
thymus"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAAGAACGCTCTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGACATTCTCGAGTAATTAATCCCCCCCCCC
3']. cDNA was cleaved with XbaI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

```

**RESULT 7**

BB638573	BB638573	458 bp mRNA linear EST 31-AUG-2001
LOCUS	BB638573	RIKEN full-length enriched, 3 days neonate thymus Mus
DEFINITION	BB638573	mRNA clone A630041P07 5', mRNA sequence.
ACCESSION	BB638573.1	GI:15401196
VERSION		
KEYWORDS	Mus musculus (house mouse)	
SOURCE	Mus musculus	
ORGANISM	Mus musculus	

**Mammalia: Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;**  
**Mammalia: Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.**  
**1 (bases 1 to 458)**

**REFERENCE**

AUTHORS	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramori,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuura,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Orido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Segabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muranatsu,M., and Hayashizaki,Y.
JOURNAL	Unpublished (2001)
COMMENT	Contact: Yoichiro Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayasi,Y., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to o prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanuki,M., Yoneda,Y., Ishikawa,T., Tanaka,T., Matsuuwa,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer. Genome Res. 10 (11), 1767-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y., and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 288-289 (2001) Yamamoto,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Azakawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y., and Hayashizaki,Y. Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Funct. Genomics 2 pre, 172-L6 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES	Location/Qualifiers
source	1 - 458 Organism="Mus musculus" /mol type="mRNA" /db_xref="taxon:10090" /clone="A630041P07" /tissue_type="thymus"

Analysis of the mouse transcriptome based on functional annotation	
of 60,770 full-length cDNAs Nature 420, 563-573 (2002)	
JOURNAL MEDLINE PUBLMED COMMENT	<p>Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.go.jp/ URL: http://genome.gsc.riken.go.jp/</p> <p>Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kegawa, I., Kawai, J., Kojima, Y., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, H., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasai, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanuki, A., Muramatsu, M. and Hayashizaki, Y.</p> <p>Direct Submission</p> <p>Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)</p> <p>Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new nonredundant cDNA library. Genome Res. 11 (10), 1617-1630 (2000)</p> <p>RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)</p> <p>Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a cDNA library. Genome Res. 11 (2), 281-289 (2001)</p> <p>cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.</p> <p>Please visit our web site (<a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a>) for further details.</p>
FEATURES SOURCE	<p>Location/Qualifiers</p> <p>1. .667 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone_id="G830026K01" /tissue_type="mammary gland" /cell_line="RCB-0526 JY9-MC(A)" /clone_lib="RIKEN full-length enriched, mammary gland RCB-0526 JY9-MC(A) cDNA"</p>
FEATURES SOURCE	<p>Query Match 17.2%; Score 45; DB 13; Length 667; Best Local Similarity 100.0%; Pred. No. 1.9e-09; Indels 0; Gaps 0;</p>
RESULT 9 LOCUS DEFINITION	<p>BUT0706646 BUT0706646-cad-P-11-0-U1.r1 mRNA linear EST 15-JUL-2003 IMAGE:6409798 5', mRNA sequence.</p>
ACCESSION KEYWORDS	<p>EST.</p>
SOURCE ORGANISM	<p>Mus musculus (house mouse)</p>
RESULT 9 LOCUS DEFINITION	<p>BUT0706646 BUT0706646 702 bp mRNA linear EST 15-JUL-2003 IMAGE:6409798 5', mRNA sequence.</p>
ACCESSION KEYWORDS	<p>EST.</p>
SOURCE ORGANISM	<p>Mus musculus (house mouse)</p>
REFERENCE AUTHORS TITLE JOURNAL COMMENT	<p>1 (bases 1 to 702) NIH-MGC http://mgc.ncbi.nlm.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov</p> <p>Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the T.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a></p> <p>This clone was contributed by the Brain Molecular Anatomy Project (BMAP).</p> <p>Seq Primer: PYX-5' Location/Qualifiers</p> <p>source 1. .702 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE:640978" /tissue_type="whole brain" /dev_stage="embryo/12.5dpc" /lab_host="DE10B (T1 phage resistant)" /clone_id="NTH_BMAP_F00" /note="Organ: Brain; Vector: PYX-5'; Site: 1; EcoRI I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoRI adaptor, digested with NotI and then cloned directionally into PYX-5' vector. The library tag sequence located between the Not I site and the polyA tail is TGAGGAGCC. This library was created for the University of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System, supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."</p> <p>ORIGIN</p> <p>Query Match 17.2%; Score 45; DB 13; Length 702; Best Local Similarity 100.0%; Pred. No. 1.9e-09; Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p> <p>Qy 1 GATCAAGGTGGAGCTTGAGGAGCTGTCGAGACAAAGCGGCCG 45 Db 310 GATCAAGGTGGAGCTTGAGGAGCTGTCGAGACAAAGCGCCG 354</p>
REFERENCE AUTHORS TITLE JOURNAL COMMENT	<p>BQ960960 BQ960960 930 bp mRNA linear EST 21-AUG-2002 AGENCourt 8955743 NCT_CGAP Mam2 Mus musculus cDNA clone IMAGE:6439792 5', mRNA sequence.</p> <p>Accession BQ960960 Version BQ960960.1 Keywords EST. Source Mus musculus (house mouse)</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 930) NIH-MGC http://mgc.ncbi.nlm.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov</p> <p>Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc.</p>

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (L1NL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/L1NL at:  
<http://image.lnl.gov>  
 Plate: L1AM13960 row: f column: 17  
 High quality sequence stop: 603.  
 Location/Qualifiers  
 1..930  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:16439792"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /clone\_id="NCT CGAP Mam2"  
 /note="Organ: mammary; Vector: pcMV-SPORT6; Site 1: Sall;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dR.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

	Query	Match	Score	DB	Length
Best Local Similarity	100.0%	Pred. No.	2e-09;		
Matches	45;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0
Y	1	GATCAAGTGGAACTGAGGACTGAGCTGTGAGACCAAAGACGGCGG	45		
jb	335	GATCAAGTGGAACTGAGGACTGAGCTGTGAGACCAAAGACGGCGG	379		

RESULT 11  
 K014847 AK041847 2052 bp mRNA linear HTC 19-SEP-2005  
 OCUS Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched  
 DEFINITION library, Clone: A63041P07 product-hypothetical PDZ domain (also  
 known as DHR or GLGF) containing protein, full insert sequence.  
 ACCESSION AK041847  
 KEYWORDS GT:6088683  
 SOURCE HTC; CAP trapper.  
 ORGANISM Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 \*\*\*\*

AUTHORS	Carninci, P. and Hayashizaki, Y.	AUTHORS	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning	TITLE	Normalization and subtraction of cap-trapper selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new gene
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)	JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	99279253	MEDLINE	10349636
PUBMED		PUBMED	2
REFERENCE		REFERENCE	

MEDLINE 2049374  
 PUBMED 1104259  
 REFERENCE 3  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nisui, K., Kitaura, T., Tashiro, H., Itch, M., Sumi, N., Ishii, Y., Nakamura, S., Harama, M., Nishimine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujimake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watohiki, M., Yoneida, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuya, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, T., and Hayashi, S., Kaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system -384-Format

JOURNAL *Sequencing pipeline with quality controllability sequences*  
Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE  
PUBMED 11076563



[http://www.genoscope.cns.fr/cg-bin/cluster.cgi?seq=CS0BAF039ZH05\\_AF03689\\_2&cluster=5483.r](http://www.genoscope.cns.fr/cg-bin/cluster.cgi?seq=CS0BAF039ZH05_AF03689_2&cluster=5483.r)

Contact : Feng Liang Email: fling@lifetech.com Com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0BAF039ZH05\_AF03689\_2.

## FEATURES

## source

1. 1.079  
   /organism="Homo sapiens"  
   /mol type="mRNA"  
   /db\_xref="taxon:9606"  
   /clone="CSODK002YA12"  
   /cell type="HELA CELLS COT 25-NORMALIZED"  
   /cell\_line="HELA"  
   /clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
   /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match Score 33; DB 13; Length 1079;  
 Best Local Similarity 100.0%; Pred. No. 0.00049;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 229 CAAGGGGCTTCACCTGGCCATCGACGTGGA 261  
   ||||| ||||| ||||| ||||| ||||| |||||  
 Db 30 CAACGGGGCTTCACCTGGCCATCGACGTGGA 62

RESULT 15  
 Locus BB655451 464 bp mRNA linear EST 26-OCT-2001  
 DEFINITION Full-length enriched, 9 days embryo Mus musculus cDNA clone D030060C17 5', mRNA sequence.  
 ACCESSION BB655451  
 VERSION BB655451.1  
 GI: 16489279  
 EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramatsu, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sugabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.

## TITLE

## JOURNAL

## COMMENT

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 Unpublished (2001)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp  
 URL:<http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
 Nagai, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)  
 Ronno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Ichinomori, K., Kawai, J., Shibata, K., and Hayashizaki, Y.  
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequence. *Mamm. Genome* 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

Location/Qualifiers

1. 464  
   /organism="Mus musculus"

/mol type="mRNA"  
   /db\_xref="taxon:10090"  
   /clones "D030060C17"  
   /dev stage="9 days embryo"  
   /lab host="DH10B"

/clone lib="RIKEN full-length enriched, 9 days embryo"  
   /note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in RIKEN contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5, GAGAGGAGACGGGCCACTGACTGTTTCTTCTTCTTCTVN 3'], cDNA was prepared by using trehalose-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5, GAGAGGAGGATCTCGATTATATATCCCCCCCCCC 3'], cDNA was cleaved with BamHI and XbaI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

## ORIGIN

Query Match Score 23; DB 10; Length 464;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 239 TTCCACCTGGCCATGAGCTGGA 261  
   ||||| ||||| ||||| ||||| |||||  
 Db 70 TTCCACCTGGCCATGAGCTGGA 92

Search completed: July 18, 2004, 15:49:14  
 Job time : 1818 secs



Result No.	Score	Query Match	Length	DB ID	Description
1	261	100.0	4303	4	US-09-976-594-899 Sequence 899, Appl
2	21	8.0	1425	2	US-08-356-060A-6 Sequence 6, Appl
3	21	8.0	1425	3	US-08-460-00C-6 Sequence 6, Appl
4	21	8.0	1425	3	US-08-674-09B-6 Sequence 6, Appl
5	21	8.0	1425	3	US-08-954-598-6 Sequence 6, Appl
6	21	8.0	1425	4	US-08-957-74-6 Sequence 6, Appl
7	21	8.0	1425	4	US-09-325-56-10 Sequence 10, Appl
8	21	8.0	1425	4	US-09-639-995-6 Sequence 6, Appl
9	21	8.0	1425	4	US-09-448-188-6 Sequence 6, Appl
10	21	8.0	1425	4	US-08-954-128-6 Sequence 6, Appl
11	21	8.0	1425	4	US-09-704-917-6 Sequence 6, Appl
12	21	8.0	1425	4	US-08-954-740-6 Sequence 6, Appl
13	21	8.0	1425	4	US-09-151-939-6 Sequence 6, Appl
14	21	8.0	1425	4	US-09-736-4176-6 Sequence 6, Appl
15	21	8.0	1576	1	US-08-748-511-5 Sequence 5, Appl
16	21	8.0	1576	1	US-08-748-591-10 Sequence 10, Appl
17	21	8.0	2274	4	US-09-772-647-3 Sequence 3, Appl
c 18	18	6.9	96	3	US-08-184-322-24 Sequence 24, Appl
c 19	18	6.9	402	2	US-08-193-078B-20 Sequence 20, Appl
c 20	18	6.9	3225	5	PCT-US53-0651-91 Sequence 45, Appl
c 21	18	6.9	3225	5	PCT-US53-0651-91 Sequence 91, Appl
c 22	17	6.5	291	4	US-09-252-991A-10114 Sequence 10114, A
c 23	17	6.5	390	4	US-09-252-991A-4800 Sequence 4800, Ap
c 24	17	6.5	552	4	US-09-252-991A-5775 Sequence 4575, Ap
c 25	17	6.5	594	4	US-09-252-991A-5966 Sequence 9966, Ap
c 26	17	6.5	648	4	US-09-252-991A-4749 Sequence 4749, Ap
c 27	17	6.5	681	4	US-09-252-991A-4694 Sequence 4694, Ap

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: July 18, 2004, 14:47:40 ; Search time 56 Seconds  
(without alignments)  
Title: US-10-051-769-2  
Perfect score: 261  
Sequence:  
1 gateaggctggatgtcgggg.....cacctggccatcgacgttggaa 261  
Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 682709 seqs, 277475446 residues  
Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgr2\_6/podata/2/ina/5A COMB. seq:  
2: /cgr2\_6/podata/2/ina/5B COMB. seq:  
3: /cgr2\_6/podata/2/ina/5A COMB. seq:  
4: /cgr2\_6/podata/2/ina/5B COMB. seq:  
5: /cgr2\_6/podata/2/ina/PCTUS COMB. seq:  
6: /cgr2\_6/podata/2/ina/backfile1.seq:  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
---	---	---	---	---	---
1	261	100.0	4303	4	US-09-976-594-899 Sequence 899, Appl
2	21	8.0	1425	2	US-08-356-060A-6 Sequence 6, Appl
3	21	8.0	1425	3	US-08-460-00C-6 Sequence 6, Appl
4	21	8.0	1425	3	US-08-674-09B-6 Sequence 6, Appl
5	21	8.0	1425	4	US-08-954-598-6 Sequence 6, Appl
6	21	8.0	1425	4	US-08-957-74-6 Sequence 6, Appl
7	21	8.0	1425	4	US-09-325-56-10 Sequence 10, Appl
8	21	8.0	1425	4	US-09-639-995-6 Sequence 6, Appl
9	21	8.0	1425	4	US-09-448-188-6 Sequence 6, Appl
10	21	8.0	1425	4	US-08-954-128-6 Sequence 6, Appl
11	21	8.0	1425	4	US-09-704-917-6 Sequence 6, Appl
12	21	8.0	1425	4	US-08-954-740-6 Sequence 6, Appl
13	21	8.0	1425	4	US-09-151-939-6 Sequence 6, Appl
14	21	8.0	1425	4	US-09-736-4176-6 Sequence 6, Appl
15	21	8.0	1576	1	US-08-748-511-5 Sequence 5, Appl
16	21	8.0	1576	1	US-08-748-591-10 Sequence 10, Appl
17	21	8.0	2274	4	US-09-772-647-3 Sequence 3, Appl
c 18	18	6.9	96	3	US-08-184-322-24 Sequence 24, Appl
c 19	18	6.9	402	2	US-08-193-078B-20 Sequence 20, Appl
c 20	18	6.9	3225	5	PCT-US53-0651-91 Sequence 45, Appl
c 21	18	6.9	3225	5	PCT-US53-0651-91 Sequence 91, Appl
c 22	17	6.5	291	4	US-09-252-991A-10114 Sequence 10114, A
c 23	17	6.5	390	4	US-09-252-991A-4800 Sequence 4800, Ap
c 24	17	6.5	552	4	US-09-252-991A-5775 Sequence 4575, Ap
c 25	17	6.5	594	4	US-09-252-991A-5966 Sequence 9966, Ap
c 26	17	6.5	648	4	US-09-252-991A-4749 Sequence 4749, Ap
c 27	17	6.5	681	4	US-09-252-991A-4694 Sequence 4694, Ap

## ALIGNMENTS

RESULT 1  
US-09-976-594-899  
; Sequence 899, Application US/09976594  
; Patent No. 6673149  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976, 594  
; CURRENT FILING DATE: 2001-10-12  
; PRIORITY APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL program  
; SEQ ID NO: 899  
; LENGTH: 4303  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1  
US-09-976-594-899

Query Match 100.0%; Score 261; DB 4; Length 4303;  
Best Local Similarity 100.0%; Pred. No. 2.5e-110; Mismatches 0; Indels 0; Gaps 0;

Matches 261; Conservative 0; MisMatches 0; Del 79 GATCAAGGTGGATGTCGAGGCTCGTGCAGACAGACGCCGGCTGAGCCATCGGGT 120  
Qy 1 GATCAAGGTGGATGTCGAGGCTCGTGCAGACAGACGCCGGCTGAGCCATCGGGT 60  
Db 79 GATCAAGGTGGATGTCGAGGCTCGTGCAGACAGACGCCGGCTGAGCCATCGGGT 138  
Qy 61 GCTGAGCCCTGGGAAGTGTGTTCCNGGCGAGTGAGCTGGAGCGAGCTGGCTGAGCCATCGGGT 120  
Db 139 GCTGAGCCCTGGGAAGTGTGTTCCNGGCGAGTGAGCTGGAGCGAGCTGGCTGAGCCATCGGGT 198  
Qy 121 CGTGCAGCCAGTCGCTGCCCTGGCCAGACGGGAGCCGGCTGAGCCATCGGGT 180  
Db 199 CGTGCAGCCAGTCGCTGCCCTGGCCAGACGGGAGCCGGCTGAGCCATCGGGT 258  
Qy 241 CCACCTGGCATCGAGCTGGA 319 CCACCTGGCATCGAGCTGGA 339

RESULT 2

CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109

GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing Proteins and Uses Related Thereto  
 NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:  
 ADDRESSEES: LAHTIVE & COCKFIELD  
 STREET: 60 State Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/460,900C  
 FILING DATE: 5-JUNE-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/435,093  
 FILING DATE: 4-DECEMBER-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/356,060  
 FILING DATE: 14-DEC-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 832-1000  
 TELEFAX: (617) 832-7400  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1425  
 US-08-460-900C-6

Query Match 8.0%; Score 21; DB 2; Length 1425;  
 Best Local Similarity 100.0%; Pred No. 0.87;  
 Matches 21; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

RESULT 4  
 US-08-460-900B-6  
 Sequence 6, Application US/08674509B  
 GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 APPLICANT: Marigo, Valeria  
 TITLE OF INVENTION: SCREENING ASSAYS FOR HEDGEHOG  
 NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEES: FOLEY, HOAG & ELLIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109-2170

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.1.30

RESULT 3  
 US-08-460-900C-6  
 Sequence 6, Application US/08460900C  
 GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 APPLICANT: Bumcrot, David A.  
 APPLICANT: Marti-Gorostiza, Elisa  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing Proteins and Uses Related Thereto  
 NUMBER OF SEQUENCES: 62  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEES: FOLEY, HOAG & ELLIOT LLP  
 STREET: One Post Office Square

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/674,509B  
 FILING DATE: 02-JUL-1996  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/460,900  
 FILING DATE: 05-JUN-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: HMV-006 .06  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-832-1000  
 TELEFAX: 617-832-1000  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1425  
 US-08-674-509B-6

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1425  
 US-08-674-509B-6

Query Match Score 21; DB 3; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 0.87%;  
 Matches 21; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

RESULT 6  
 US-08-957-874-6

Sequence 6, Application US/08957874  
 GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 Patent No. 638A192  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing  
 NUMBER OF SEQUENCES: 47  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/954,698  
 FILING DATE: 20-OCT-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/462,386  
 FILING DATE: 05-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/435,093  
 FILING DATE: 04-MAY-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/356,060  
 FILING DATE: 14-DEC-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: HMV-006 .09  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 832-1000  
 TELEFAX: (617) 832-7000

Query Match Score 21; DB 3; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 0.87%;  
 Matches 21; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

RESULT 5  
 US-08-554-698-6

Sequence 5, Application US/08954698  
 GENERAL INFORMATION:  
 PATENT NO. 6271363  
 APPLICANT: Ingham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing  
 NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109-2170  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/954,698  
 FILING DATE: 20-OCT-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/462,386  
 FILING DATE: 20-OCT-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/435,093  
 FILING DATE: 04-MAY-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/356,060  
 FILING DATE: 14-DEC-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: HMV-006 .09  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 832-1000  
 TELEFAX: (617) 832-7000

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1425

US-08-957-874-6

RESULT 7

Query Match Score 21; DB 4; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 0.87;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 139 CTGGCCACCGGGAGGCCGA 159  
 Db 837 CTGGCCACGGAGGCCGA 857

US-09-325-256-10

Sequence 10, Application US/09325256

Patent No. 6444793

GENERAL INFORMATION:  
 APPLICANT: PEPINSKY, R. BLAKE  
 BAKER, DARREN P.  
 WEN, DINGYI  
 APPLICANT: WILLIAMS, KEVIN P.  
 GARGER, ELLEN A.  
 APPLICANT: TAYLOR, FREDERICK R.  
 GALDES, ALPHONSE  
 APPLICANT: PORTER, JEFFREY  
 TITLE OF INVENTION: HYDROPHOBICALLY-MODIFIED PROTEIN COMPOSITIONS AND METHODS

FILE REFERENCE: BIV-067.01

CURRENT FILING DATE: 1999-06-03  
 CURRENT APPLICATION NUMBER: US/09/325,256

PRIOR APPLICATION NUMBER: 60/099,800  
 PRIOR FILING DATE: 1998-09-10  
 PRIOR APPLICATION NUMBER: 60/078,935  
 PRIOR FILING DATE: 1998-03-20  
 PRIOR APPLICATION NUMBER: 60/089,685  
 PRIOR FILING DATE: 1998-06-17  
 PRIOR APPLICATION NUMBER: 60/067,423  
 PRIOR FILING DATE: 1997-12-03  
 PRIOR APPLICATION NUMBER: PCT/US98/25676  
 PRIOR FILING DATE: 1998-12-03  
 NUMBER OF SEQ ID NOS: 31  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 10  
 LENGTH: 1425  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: n = a, t, g or c  
 LOCATION: (1)..(1425)  
 US-09-325-256-10

Query Match Score 21; DB 4; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 0.87;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 139 CTGGCCACCGGGAGGCCGA 159  
 Db 837 CTGGCCACGGAGGCCGA 857

RESULT 8

US-09-639-695-6

; Sequence 6, Application US/09639695

Patent No. 6576237

GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 McMahon, Andrew P.  
 Tabin, Clifford J.  
 Bumcrot, David A.  
 Marti-Gorostiza, Elissa  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing Proteins and Uses Related Thereto

NUMBER OF SEQUENCES: 62  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/639,695  
 FILING DATE: 16-Aug-2000  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US/09/639,695  
 FILING DATE: 4-May-1995  
 APPLICATION NUMBER: US 08/356,060  
 FILING DATE: 14-DEC-1994  
 APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36709  
 REFERENCE/DOCKET NUMBER: HMV-006-005  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 832-1000  
 TELEFAX: (617) 832-7000  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1425  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 US-09-639-695-6

Query Match Score 21; DB 4; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 0.87;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 139 CTGGCCACCGGGAGGCCGA 159  
 Db 837 CTGGCCACGGAGGCCGA 857

RESULT 9

US-09-448-188-6

; Sequence 6, Application US/09448188

GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 McMahon, Andrew P.  
 Tabin, Clifford J.  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing Proteins and Uses Related Thereto

NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109-2170

COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent/PC Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/448,188  
 FILING DATE: 23-NO- 6607913-1999

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/462,386  
 FILING DATE: 05-JUN-1995  
 APPLICATION NUMBER: US 08/435,093  
 FILING DATE: 04-MAY-1995  
 APPLICATION NUMBER: US 08/356,060  
 FILING DATE: 14-DEC-1994  
 APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993

ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: HMV-006.12

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-832-1000  
 TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1435 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:

NAME/KEY:	CDS
LOCATION:	1..1425
SEQUENCE DESCRIPTION:	SEQ ID NO: 6:

S-09-448-188-6

Query Match 8.0%; Score 21; DB 4; Length 14  
 Best Local Similarity 100.0%; Pred. No. 0.87;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0

Query	139 CTGGCCACCGGGAGGCCGA 159
DB	837 CTGGCCACCGGGAGGCCGA 857

RESULT 10  
 S-08-954-128-6

Sequence 6, Application US/08954128  
 Patent No. 6610656

GENERAL INFORMATION:  
 APPLICANT: Inigham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Induc-  
 TITLE OF INVENTION: Proteins and Uses Related Thereto  
 NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109-2170

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/954,128  
 FILING DATE: 20-OCT-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/462,386  
 FILING DATE: 05-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/435,093  
 FILING DATE: 04-MAY-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/356,060  
 FILING DATE: 14-DEC-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE DOCKET NUMBER: HNV-006.12  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-832-1000  
 TELEFAX: 617-832-1000  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1425

US-08-954-128-6

Query Match 8.0%; Score 21; DB 4; Length 1425  
 Best Local Similarity 100%; Pred. No. 0.87;  
 Matches 21; Conservative 0; Mismatches 0;

Qy 139 CTGGGCCACGGGAGGCCGA 159  
 Db 837 CTGGCCACGGGAGGCCGA 857

RESULT 11  
 US-09-704-917-6  
 Sequence 6, Application US/09704917  
 Patent No. 616926  
 GENERAL INFORMATION:  
 APPLICANT: Biogen, Inc.  
 APPLICANT: Burkly, Linda  
 APPLICANT: Wang, Li  
 APPLICANT: Chun  
 TITLE OF INVENTION: METHOD OF MODULATING LIPID METABOLISM  
 FILE REFERENCE: A069PCT  
 CURRENT APPLICATION NUMBER: US/09/704,917  
 CURRENT FILING DATE: 2000-11-02  
 PRIOR APPLICATION NUMBER: 60/122,640  
 PRIOR FILING DATE: 1999-03-03  
 PRIOR APPLICATION NUMBER: 60/124,446  
 PRIOR FILING DATE: 1999-03-15  
 NUMBER OF SEQ ID NOS: 22  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 6  
 LENGTH: 1425  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(1425)  
 OTHER INFORMATION: "mn" encoding "Xaa" at position 9, other or unknown

US-09-704-917-6

Query Match 8.0%; Score 21; DB 4; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 0.87%;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 837 CTCGGCCACCGGGAGGCCGA 857  
 RESULT 13  
 US-09-151-999-6

Qy 139 CTCGGCCACCGGGAGGCCGA 159  
 Db 837 CTCGGCCACCGGGAGGCCGA 857

GENERAL INFORMATION:  
 APPLICANT: Ingham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing Proteins and Uses Related Thereto  
 NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: POLEY, HOAG & ELIOT LLP  
 STREET: One Post Office Square  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109-2170  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US 08/462,386  
 FILING DATE: 20-OCT-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/462,386  
 FILING DATE: 05-JUN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/435,093  
 FILING DATE: 04-MAY-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/356,060  
 FILING DATE: 14-DEC-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-832-1000  
 TELEFAX: 617-832-7000  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: both  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 LOCATION: 1..1425  
 NAME/KEY: CDS  
 Db 837 CTCGGCCACCGGGAGGCCGA 857  
 RESULT 13  
 US-09-151-999-6

Qy 139 CTCGGCCACCGGGAGGCCGA 159  
 Db 837 CTCGGCCACCGGGAGGCCGA 857

GENERAL INFORMATION:  
 APPLICANT: Wang, Elizabeth  
 TITLE OF INVENTION: REGULATION OF EPITHELIAL TISSUE BY HEDGEHOG-LIKE POLYPEPTIDES, AND FORMULATIONS AND USES RELATED THERETO  
 FILE REFERENCE: ONV-031.02  
 CURRENT APPLICATION NUMBER: US/09/151,999  
 CURRENT FILING DATE: 1998-08-11  
 EARLIER APPLICATION NUMBER: 08/955,552  
 EARLIER FILING DATE: 1997-10-20  
 NUMBER OF SEQ ID NOS: 28  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 6  
 LENGTH: 1425  
 TYPE: DNA  
 ORGANISM: Homo sapiens Shh  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1) .. (1425)  
 FEATURE:  
 OTHER INFORMATION: "nnnn" encoding "Xaa" at position 137-1389 may be a, t, c,  
 US-09-151-999-6

Query Match 8.0%; Score 21; DB 4; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 0.87%;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 837 CTCGGCCACCGGGAGGCCGA 857  
 RESULT 13  
 US-09-151-999-6

Qy 139 CTCGGCCACCGGGAGGCCGA 159  
 Db 837 CTCGGCCACCGGGAGGCCGA 857

GENERAL INFORMATION:  
 APPLICANT: Ingaham, Phillip W.  
 APPLICANT: McMahon, Andrew P.  
 APPLICANT: Tabin, Clifford J.  
 APPLICANT: Bumcrot, David A.  
 APPLICANT: Marti-Gorostiza, Elisa  
 TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing Proteins and Uses Related Thereto  
 NUMBER OF SEQUENCES: 47  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHVE & COCKFIELD  
 STREET: 60 State Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII (text)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/736,476  
 FILING DATE: 13-Dec-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 4 MAY-1995  
 FILING DATE: 4 MAY-1995  
 APPLICATION NUMBER: US 08/435,093  
 FILING DATE: 14-DEC-1994

APPLICATION NUMBER: US 08/176,427  
 FILING DATE: 30-DEC-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Vincent, Matthew P.  
 REGISTRATION NUMBER: 36,709  
 REFERENCE/DOCKET NUMBER: RMT-006CP4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1425 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1425  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 Us-09-736-476-6

Query Match 8.0%; Score 21; DB 1; Length 1576;  
 Best Local Similarity 100.0%; Pred. No. 0.86;  
 Matches 21; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

QY 139 CTCGGCCACGGGAGGCCGA 159  
 Db 988 CTCGGCCACGGGAGGCCGA 1008

Search completed: July 18, 2004, 15:54:55  
 Job time : 58 secs

---

RESULT 15  
 US-08-48-591-5  
 Sequence 5, Application US/08748591  
 Patent No. 5759811  
 GENERAL INFORMATION:  
 APPLICANT: Epstein, Ervin  
 APPLICANT: Hu, Zhilan  
 APPLICANT: Bonitas, Jeanette  
 TITLE OF INVENTION: Mutant Human Hedgehog Gene  
 NUMBER OF SEQUENCES: 23  
 ADDRESSEE: Fish and Richardson  
 STREET: 2200 Sand Hill Road  
 CITY: Menlo Park  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94025  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/748,591  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sherwood, Pamela J.  
 REGISTRATION NUMBER: 36,677  
 REFERENCE/DOCKET NUMBER: 06510/067001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 854-0875  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1576 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-08-748-591-5



Copyright GenCore version 5.1.6  
(c) 1993 - 2004 Compugen Ltd.

NM nucleic search, using sw model

Run on: July 18, 2004, 12:32:20 ; Search time 88.141 Seconds  
(without alignments)

Scoring table: IDENTITY NUC GapOp 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:  
5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq:  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:  
\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

result	No.	Score	Query	Match	Length	DB	ID	Description
1	261	100.0	4303	4	US-09-976-594-899	Sequence 899, App		Sequence 899, App
2	51.6	19.8	6436	4	US-09-976-599-1	Sequence 1, Appli		Sequence 1, Appli
3	44.4	17.0	450	4	US-09-252-991A-8011	Sequence 8011, Ap		Sequence 8011, Ap
4	44.4	17.0	1467	4	US-9-251-991A-8188	Sequence 8188, Ap		Sequence 8188, Ap
5	44.4	17.0	2823	4	US-09-252-991A-0068	Sequence 8068, Ap		Sequence 8068, Ap
6	44.4	17.0	2949	4	US-09-252-991A-0249	Sequence 8249, Ap		Sequence 8249, Ap
7	44.4	17.0	23673	4	US-09-773-816-1	Sequence 1, Appli		Sequence 1, Appli
8	43.6	16.6	4849	4	US-09-220-723-39	Sequence 39, Appli		Sequence 39, Appli
9	43.4	16.6	403765	3	US-09-103-840-A-2	Sequence 2, Appli		Sequence 2, Appli
10	43.4	16.6	411529	3	US-09-103-840-A-1	Sequence 1, Appli		Sequence 1, Appli
11	43.2	16.6	2825	4	US-09-196-390-5	Sequence 1, Appli		Sequence 1, Appli
12	42.4	16.2	77536	4	US-09-410-512-B-1	Sequence 1, Appli		Sequence 1, Appli
13	42.2	16.2	450	4	US-09-252-991A-10221	Sequence 10221, A		Sequence 10221, A
14	42.2	16.2	561	4	US-09-252-991A-10438	Sequence 10438, A		Sequence 10438, A
15	42.2	16.2	705	4	US-09-252-991A-10340	Sequence 10340, A		Sequence 10340, A
16	42.2	16.2	1053	4	US-09-252-991A-13541	Sequence 13541, A		Sequence 13541, A
17	42.2	16.2	1167	4	US-09-152-991A-10810	Sequence 10810, A		Sequence 10810, A
18	42.2	16.2	1557	4	US-09-252-991A-13907	Sequence 13907, A		Sequence 13907, A
19	42.2	16.2	1722	4	US-09-252-991A-13007	Sequence 13007, A		Sequence 13007, A
20	41.8	16.0	3917	4	US-10-164-555-7	Sequence 7, Appli		Sequence 7, Appli
21	41.6	15.9	1910	3	US-09-593-711A-3	Sequence 3, Appli		Sequence 3, Appli
22	41.6	15.9	1914	1	US-07-601-094-1	Sequence 1, Appli		Sequence 1, Appli
23	41.6	15.9	1914	1	US-08-012-725-1	Sequence 1, Appli		Sequence 1, Appli
24	41.4	15.9	1194	4	US-09-252-991A-11352	Sequence 11352, A		Sequence 11352, A
25	41.4	15.9	1500	4	US-09-252-991A-11312	Sequence 11312, A		Sequence 11312, A
26	41.4	15.7	2618	4	US-09-557-991A-14375	Sequence 25, Appli		Sequence 25, Appli
27	41.4	15.7	1398	4	US-09-232-991A-14375	Sequence 5563, Ap		Sequence 5563, Ap

**ALIGNMENTS**

RESULT 1  
US-09-976-594-899  
; Sequence 899, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Buchbinder, Jerry  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH S-  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIORITY APPLICATION NUMBER: 60/244,409  
; PRIORITY FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SEQ ID NO: 899  
; LENGTH: 893  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1  
US-09-976-594-899

Query Match 100.0%; Score 261; DB 4; Length 4303;  
Best Local Similarity 100.0%; Pred. No. 5; 5e-48;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAGGTGGAGTGGAGGAGCTGTCAGACCAAGAACGGCGCCGCGCTGGAGGG 60  
Db 79 GATCAGGTGGAGTGGAGGAGCTGTCAGACCAAGAACGGCGCCGCGCTGGAGGG 138  
QY 61 GCTGACCCCTGGGGAGCTGGTGCCTGGGGAGCCTGGCCCTCATOAAGCCATCGGGCT 120  
Db 139 GCTGACCCCTGGGGAGCTGGTGCCTGGGGAGCAGCTGGCCCTCATOAAGCCATCGGGCT 198  
QY 121 CGTCGCCGCGCCACTCGTCCTGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGCT 180  
Db 199 CGTCGCCGCGCCACTGGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGCT 258  
QY 181 GCCCGCCGCGCCCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGCT 240  
Db 259 GCCCGCCGCGCCACTGGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGCT 318

RESULT 2  
US-09-976-594-899  
; Sequence 899, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH S-  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIORITY APPLICATION NUMBER: 60/244,409  
; PRIORITY FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SEQ ID NO: 899  
; LENGTH: 893  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1  
US-09-976-594-899

Query Match 100.0%; Score 261; DB 4; Length 4303;  
Best Local Similarity 100.0%; Pred. No. 5; 5e-48;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAGGTGGAGTGGAGGAGCTGTCAGACCAAGAACGGCGCCGCGCTGGAGGG 60  
Db 79 GATCAGGTGGAGTGGAGGAGCTGTCAGACCAAGAACGGCGCCGCGCTGGAGGG 138  
QY 61 GCTGACCCCTGGGGAGCTGGTGCCTGGGGAGCCTGGCCCTCATOAAGCCATCGGGCT 120  
Db 139 GCTGACCCCTGGGGAGCTGGTGCCTGGGGAGCAGCTGGCCCTCATOAAGCCATCGGGCT 198  
QY 121 CGTCGCCGCGCCACTCGTCCTGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGCT 180  
Db 199 CGTCGCCGCGCCACTGGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGCT 258  
QY 181 GCCCGCCGCGCCCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGCT 240  
Db 259 GCCCGCCGCGCCACTGGGGAGCTGGGGAGCTGGGGAGCTGGGGAGCTGGGGCT 318

US-09-600-099-1 Sequence 1, Application US/09600099  
 : Patient No. 6649382  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEE, Sang Yoo  
 ; APPLICANT: CHOI, Jong-il  
 ; APPLICANT: CHOO, Seung-Ho  
 ; APPLICANT: YOON, Hye-Sung  
 ; APPLICANT: HAN, Kyuboem  
 ; APPLICANT: SONG, Ji-Tong  
 ; APPLICANT: LEE, Yong-Hyun  
 ; APPLICANT: HUH, Tae-Lim  
 ; APPLICANT: HONG, Sung-Kook  
 ; TITLE OF INVENTION: POLYHYDROXYALKANOATE BIOSYNTHESIS-RELATED GENES DERIVED FROM Alcaligenes latus  
 ; TITLE OF INVENTION: Ligenes latus  
 ; FILE REFERENCE: 428-1001  
 ; CURRENT APPLICATION NUMBER: US/09/600,099  
 ; CURRENT FILING DATE: 2000-07-11  
 ; PRIOR APPLICATION NUMBER: KR 98-1422  
 ; PRIOR FILING DATE: 1998-01-19  
 ; PRIOR APPLICATION NUMBER: KR 98-1423  
 ; PRIOR FILING DATE: 1998-01-19  
 ; PRIOR APPLICATION NUMBER: KR 98-58760  
 ; PRIOR FILING DATE: 1998-12-26  
 ; PRIOR APPLICATION NUMBER: PCT/KR99/000931  
 ; PRIOR FILING DATE: 1999-01-19  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: KOPATIN 1.5  
 ; SEQ ID NO: 1  
 ; LENGTH: 6436  
 ; TYPE: DNA  
 ; ORGANISM: Alcaligenes latus  
 ; FEATURE:  
 ; NAME/KEY: gene  
 ; LOCATION: (1). (6436)  
 ; OTHER INFORMATION: double stranded linear oligonucleotide for polyhydroxyalkanoate b  
 ; OTHER INFORMATION: biosynthesis-related Genes  
 US-09-600-099-1

	Query Match	Score	Length
QY	1 GATCAGGTGGAGTTCGAGGAAGCTGGCAGACCAAGACGGCGGGCGCTGCCTGAGGG	51.6;	6436;
Ddb	256 GAAGAAGGAGAAGTTCCGGAGCTATGCCGCTGATGCGCCGCTGCGCTGCGCT	50	
Qy	61 GCTGAGCTGCGGGAGCTGTCCTGGAGGGCTTGAGGCTGAGGCTATCAGACCAT-----	51.6;	114
Ddb	316 GCTGCCGCGACGGCACTGCTGCGCTGAGGCACTGATGCCATATCGAGTACCTGGACCA	375	
Qy	115 ---CCGGCTCGGGCCAGTCGCTGCCCTCGGCCACCCGGAGCCGATGCCCTGAACG	51.6;	171
Ddb	376 GACCCATTCCGAGGCCGCGCTGTGCCCCCTGGAGGCTGGCCCTTCGAGGGAGGTGAA	51.6;	435
Qy	172 GGAGGGCGCTGCGCCGCGCTGCCGAGGCTGGCCCTTCGAGGGAGGTGAA	51.6;	231
Ddb	436 GCTGGGGCAAGAACATCCCTGGAGATCACCGCGTCAACAAACCTGGCGTGTGCGCTA	51.6;	495
Qy	232 CGGGGCGCTTCACCTGCCATGAGCTGA	51.6;	261
Ddb	496 CCTGGGGCAACCTCAAGGTGGCGAGGA	51.6;	525

RESULT 3  
 US-09-252-991A-8011/C  
 ; Sequence 8011, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfeld et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196-136

**RESULT 5**  
 US-09-252-991A-8068/C  
; Sequence 8068, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196-136  
; CURRENT APPLICATION NUMBER: US/09/252-991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO: 8068  
; LENGTH: 2823  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8068

Query Match Score 44.4; DB 4; Length 2823;  
 Best Local Similarity 49.6%; Pred. No. 0.22;  
 Matches 114; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy	4	CAAGGTGAGTCAGAGCTGCTGAGACCAAGACGCCCGCCGATGCCATCG 575
Db	2142	CAATGCAGCTTACGCCAACCTGCGGACTGGGATGCGATGGATCG 575
Qy	64	GAGCCTCGGAGCTCTGGAGCTGCTGCTGGAGACGCTGCCATCAAGACCATCGCTG 123
Db	2202	GCGGAGCCGCGGAGCTGCTGCTGGAGACGCTGCCATCAAGACCATCGCTG 123
Qy	124	GGGGCGCTGCTGCTGGGACCCGTTGCTGAGGGAGGGCTGCGCCATCGAGAGCAAGGAGCTG 183
Db	2262	GATCCCCGGCATGCTGCTAGCTGACCCGGCTGATCCCCTCATCGCGCTG 183
Qy	184	CGCCGCTGCCGAAGAGCTGCCATTGAGGGAGGTGAGTAACAG 233
Db	2322	ATTCGAGGCCAGGCTGCTGCGCGATGGCAGATCG 233

**RESULT 7**  
 US-09-773-816-1  
; Sequence 1, Application US/09773816  
; Patent No. 6340774  
; GENERAL INFORMATION:  
; APPLICANT: Stanford University  
; APPLICANT: Khosla, Chaitan  
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR  
; FILE REFERENCE: 26500-20210/00  
; CURRENT APPLICATION NUMBER: US/09/773,816  
; CURRENT FILING DATE: 2001-06-08  
; PRIOR APPLICATION NUMBER: US 60/243,458  
; PRIOR FILING DATE: 2000-10-25  
; PRIOR APPLICATION NUMBER: US 60/179,305  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 1  
; LENGTH: 23673  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE: misc\_feature  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(23623)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-773-816-1

Query Match Score 44.4; DB 4; Length 23673;  
 Best Local Similarity 49.2%; Pred. No. 0.25;  
 Matches 117; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Qy	24	TCTGCGACCAAGACGCCCGGGAGCTGCTGAGGGCTGAGCCCTGAGCTG 83
Db	10394	TCCGGCGACCGAGAGTCGAGATCTCGAGGAGGCTCAGGTGCTGCC 83
Qy	84	TGGGCGAGACGCTGCCATCAAGACCATCGGCTGCTGAGGGCTGAGCTG 143
Db	11054	TGACGNGGACCTCGGGGAGATAACCTGAACTCGGCTGAGCTG 143

**RESULT 6**  
 US-09-252-991A-8249  
; Sequence 8249, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196-136  
; CURRENT APPLICATION NUMBER: US/09/252-991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO: 8249  
; LENGTH: 2949  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8249

Query Match Score 44.4; DB 4; Length 2949;  
 Best Local Similarity 49.6%; Pred. No. 0.22;  
 Matches 114; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy	144	CCACCGGGAGCCGATGGCTGAGGGAGGGCTGCTGCGCCATCGAGCTG 203
Db	11114	CCCGCCCGCTCATCGCTTACGGGACTGGAGACGGCCGATCGAGCTG 203
Qy	204	TGGCCTTCGAGGGAGGTGAGTAACCGGGCTCCACCTGGCATCGCTG 261
Db	11174	TGGGCTTCGAGGGAGGTGAGTAACCGGGCTCCACCTGGCATCGCTG 261

**RESULT 8**  
 US-09-620-312D-39  
; Sequence 39, Application US/09620312D  
; Patent No. 6559662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod  
 APPLICANT: Zhang, Jie  
 APPLICANT: Ren, Feiyian  
 APPLICANT: Chen, Rui-hong  
 APPLICANT: Zhao, Qing A.  
 APPLICANT: Wehrman, Tom  
 APPLICANT: Xue, Aidong J.  
 APPLICANT: Yang, Yonghong  
 APPLICANT: Wang, Jian-Rui  
 APPLICANT: Zhou, Ping  
 APPLICANT: Ma, Yunding  
 APPLICANT: Wang, Dunrui  
 APPLICANT: Wang, Zhiwei  
 APPLICANT: John Tillinghast  
 APPLICANT: Dramatac, Radio T.  
 TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and  
 TITLE OF INVENTION: Polypeptides  
 FILE REFERENCE: 78CIP2B  
 CURRENT APPLICATION NUMBER: US/09/620,312D  
 CURRENT FILING DATE: 2000-07-19  
 PRIOR APPLICATION NUMBER: 09/552,317  
 PRIOR FILING DATE: 2000-04-15  
 PRIOR APPLICATION NUMBER: 09/488,725  
 PRIOR FILING DATE: 2000-01-21  
 NUMBER OF SEQ ID NOS: 1105  
 SOFTWARE: pt\_FL\_genes Version 1.0  
 LENGTH: 4849  
 SEQ ID NO: 39  
 FEATURE: DNA  
 ORGANISM: Homo sapiens  
 LOCATION: (355) .. (3744)  
 US-09-620-312D-39

Query Match 16.6%; Score 43.4; DB 3; Length 4403765;  
 Best Local Similarity 51.9%; Pred. No. 0.57; Indels 0; Gaps 0;  
 Matches 98; Conservative 0; Mismatches 91; Delins 0;

Qy 38 ACGGCCGGCCCTGTGAGGGCTGAGCTGCGGACGAGCTG 97  
 Db 602195 ACGGTGGCCCGTGTGCACTGCGCAATGGCCAGCGCGCTGTCGCTGGAAAGCGAGTG 602254

Qy 98 CCCTCATAGAACATCCGCCTGGGCAAGTGTGCCACGGGAGGCC 157  
 Db 602255 CACTCGAAACCGCATTGACGCTGCGCTGCGCTGGATG 602314

Qy 158 GATGGCCCTGAAGGGAGGGCTGCGCCGCGCTGCCCGAGGAATGGCTTCGAGGCC 39  
 Db 602315 GCGAGCGMAATTGGCTCTGGGCAAGGCGCTGATCGGCCGGCG 602374

Qy 218 GAGGTGAG 226  
 Db 602375 ATGGGGCG 602383

RESULT 10 US-09-103 840A-1  
 Sequence 1, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; WHITE, Owen R.  
 ; ATTORNEY: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; FILE REFERENCE: 24366 - 2007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103 , 840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4411529  
 ; Query Match 16.6%; Score 43.4; DB 3; Length 4411529;  
 ; Best Local Similarity 51.9%; Pred. No. 0.57; Indels 0; Gaps 0;  
 ; Matches 98; Conservative 0; Mismatches 91; Delins 0;

Qy 38 ACGGGGGGGCCCTGTGAGGGCTGAGCTGCGGACGAGCTGCGGAGGCC 97  
 Db 600835 ACGGTGGCCCGTGTGCACTGCGCAATGGCCACGGGAGGCC 600894

Qy 98 CCCTCATAGAACATCCGCCTGGGCAAGTGTGCCACGGGAGGCC 157  
 Db 600895 CACTCGAAACCGCATTGACGCTGCGCTGGATG 600954

Qy 158 GATGGCCCTGAAGGGAGGGCTGCGCCGCGCTGCCCGAGGAATGGCTTCGAGGCC 217  
 Db 600955 GCGAGCGMAATTGGCTCTGGGCAAGGCGCTGATCGGCCGGCG 601014

RESULT 9 US-09-103 840A-2  
 Sequence 2, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; WHITE, Owen R.  
 ; ATTORNEY: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366 - 2007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103 , 840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4411529  
 ; Query Match 16.6%; Score 43.4; DB 3; Length 4411529;  
 ; Best Local Similarity 51.9%; Pred. No. 0.57; Indels 0; Gaps 0;  
 ; Matches 98; Conservative 0; Mismatches 91; Delins 0;

Qy 38 ACGGGGGGGCCCTGTGAGGGCTGAGCTGCGGACGAGCTGCGGAGGCC 97  
 Db 600835 ACGGTGGCCCGTGTGCACTGCGCAATGGCCACGGGAGGCC 600894

Qy 98 CCCTCATAGAACATCCGCCTGGGCAAGTGTGCCACGGGAGGCC 157  
 Db 600895 CACTCGAAACCGCATTGACGCTGCGCTGGATG 600954

Qy 158 GATGGCCCTGAAGGGAGGGCTGCGCCGCGCTGCCCGAGGAATGGCTTCGAGGCC 217  
 Db 600955 GCGAGCGMAATTGGCTCTGGGCAAGGCGCTGATCGGCCGGCG 601014







sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into Glioblastoma Multiforme, and a kit for use in detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into Glioblastoma Multiforme (a type of brain cancer). The ES is useful as molecular marker for tumour cell identification and classification and for diagnosing or identifying candidates at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is the human GLI1EN expressed sequence tag

Q	P	Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;
	Query Match	100.0% ; Score 261; DB 7; Length 261;
	Best Local Similarity	100.0% ; Pred. No. 1.2e-111;
	Matches 261;	Mismatches 0; Indels 0; Caps 0;
ay	1	GATCAAGGGAGTTCGAGGTCTGAGACCAAGA
ay	61	GCTGAGCCCTGGGAGCTGTTCTCGCGAGACGTTGCCCCCTCATTAAGACCATCCGGCT
ib	1	GATCAAGGGAGTTCGAGGTCTGAGACCAAGA
ib	61	GCTGAGCCCTGGGAGCTGTTCTCGCGAGACGTTGCCCCCTCATTAAGACCATCCGGCT
ib	61	GCTGAGCCCTGGGAGCTGTTCTCGCGAGACGTTGCCCCCTCATTAAGACCATCCGGCT
ib	121	CCTGGGGCAAGTCTGCCCTGGCCAACCGGGAGGCCGATGCCCCCTGAAGGGGGCGCT
ib	121	CCTGGGGCAAGTCTGCCCTGGCCAACCGGGAGGCCGATGCCCCCTGAAGGGGGCGCT
ib	181	GCCGGCCGGCTGCCCGAGGCTGGCCCTTGAGGGAGGTGAGACTAACGGGGCGCT
ib	181	GCCGGCCGGCTGCCCGAGGCTGGCCCTTGAGGGAGGTGAGACTAACGGGGCGCT
dy	241	CCACCTGGCATCACGTGA 261
dy	241	CCACCTGGCATCACGTGA 261

RESULT 2  
 AAD61560 standard; cDNA; 261 BP.  
 AAD61560  
 Human GLIEN EST clone cDNA #1.  
 Human glioblastoma multiforme; GBM; GLIEN; brain cancer; diagnosis;  
 gene therapy; expressed sequence tag; ss.  
*Homo sapiens.*  
 US2003108915-A1.  
 12-JUN-2003.  
 20-AUG-2002; 2002US-00224624.  
 20-OCT-2000; 2000US-0242160P.  
 20-OCT-2001; 2001US-00051769.  
 (UYN-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 McKinnon RD;  
 WPI; 2003-810848/76.  
 Novel glioblastoma multiforme associated protein GLIEN useful for  
 treating glioblastoma multiforme and diagnosing brain cancer.

Claim 38; Page 11; Opp; English.

The present invention provides novel associated protein GLIVEN useful for diagnosing brain cancer. The invention and diagnosis. The inventors sequence is human GLIVEN used in the exemplification of the invention.

Sequence 261 BP; 40 A; 81 C; 102 Ssuey Match 100.0%;

PT Glioblastoma Multiforme, designated as GLITEN, useful as a molecular marker for tumor cell identification and classification.

PT Disclosure; Page 7; 11pp; English.

XX  
 XX The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLITEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into Glioblastoma Multiforme), and a kit for use in detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into Glioblastoma Multiforme (a type of brain cancer). The EST useful as molecular marker for tumour cell identification and classification and for diagnosing or identifying candidates at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is the open reading frame from the GLITEN gene

XX Sequence 873 BP; 137 A; 304 C; 279 G; 153 T; 0 U; 0 Other;

Query Match Score 100.0%; Score 261; DB 7; Length 873;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-111; Indels 0; Gaps 0;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGTGAGTTGAGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 60  
 Db 366 GATCAAGTGAGTTGAGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 425  
 Qy 61 GCTGAGCTGGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 120  
 Db 426 GCTGAGCTGGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 485  
 Qy 121 CCTGGCCCAAGTCGTGTGCCTGGCCACGGGGATGCCCTGAAGGGAGGGCT 180  
 Db 486 CGGGGGCAAGTCGTGCCTGGCCACGGGGATGCCCTGAAGGGAGGGCT 545  
 Qy 181 GCGCGCGCTGGCCCAAGGGAGCTGGCCTGGGGCTGAGTACAAGGGGGCTT 240  
 Db 546 GCGCGCGCTGGCCCAAGGGAGCTGGCCTGGGGCTGAGTACAAGGGGGCTT 605

Qy 241 CCACCTGGCCATGAGCTTGA 261  
 Db 606 CCACCTGGCCATGAGCTTGA 626

#### RESULT 4

ABX95678

ID ABX95678 standard; DNA; 960 BP.

XX

AC ABX95678;

XX DT 30-JUN-2003 (first entry)

DE Human gene encoding GLITEN.

XX KW Human; ds; gene; chromosome 10q25; glioblastoma multiforme; brain cancer;

KW GLITEN; GMB.

XX OS Homo sapiens.

XX PH Key

FT CDS

PT /\*tag= a  
 /product= "GLITEN"

PS US2003044811-A1.

XX PD 05-MAR-2003.

XX PR 20-OCT-2001; 2001US-00051769.

XX PR 20-OCT-2000; 2000US-0242160P.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 XX McKinnon RD;  
 XX DR WPI: 2003-418995/39.

XX New isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLITEN, useful as a molecular marker for tumor cell identification and classification.  
 XX Example 2; Fig 1; 11pp; English.

XX The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLITEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into Glioblastoma Multiforme), and a kit for use in detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into Glioblastoma Multiforme (a type of brain cancer). The EST useful as molecular marker for tumour cell identification and classification and for diagnosing or identifying candidates at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is the open reading frame from the GLITEN gene

XX Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;

XX Query Match Score 100.0%; Score 261; DB 7; Length 960;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-111; Indels 0; Gaps 0;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGTGAGTTGAGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 60  
 Db 396 GATCAAGTGAGTTGAGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 455

Qy 61 GCTGAGCTGGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 120  
 Db 456 GCTGAGCTGGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 515

Qy 121 CCTGGCCCAAGTCGTGTGCCTGGCCACGGGGATGCCCTGAAGGGAGGGCT 180  
 Db 516 GCTGAGCTGGGAGCTGGTGTCTGGGAGACCCAGAACAGCGCGGCCCTGCTGGGG 575

Qy 181 GCGCGCGCTGGCCCAAGGGAGCTGGCCTGGGGCTGAGTACAAGGGGGCTT 240  
 Db 576 GCGCGCGCTGGCCCAAGGGAGCTGGCCTGGGGCTGAGTACAAGGGGGCTT 635

Qy 241 CCACCTGGCCATGAGCTTGA 261  
 Db 636 CCACCTGGCCATGAGCTTGA 656

RESULT 5		ADE07789 standard; cDNA; 960 BP.	XX	ID ADE07789 standard; DNA; 1092 BP.
ADD61559	ID AAD61559		XX	AC ADE07789;
XX	XX	DT 29-JAN-2004 (first entry)	XX	DT 29-JAN-2004 (first entry)
AC AAD61559;	XX	DE Novel coding sequence (useful for identifying genetic disorders) #855.	XX	DE Novel coding sequence (useful for identifying genetic disorders) #855.
XX	XX	XX novel gene; novel protein; tissue marker; molecular weight marker;	XX	XX novel gene; novel protein; tissue marker; molecular weight marker;
DE Human Gliten partial cDNA.	XX	XX chromosome marker; genetic disorder; gene; ds.	XX	XX chromosome marker; genetic disorder; gene; ds.
XX	XX	XX Unidentified.	OS	XX Unidentified.
KW Human; glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;	XX	XX WO2003054152-A2.	PN	XX WO2003054152-A2.
KW gene therapy; chromosome 10; gene; ss.	OS	XX PD 03-JUL-2003.	XX	XX PD 03-JUL-2003.
OS Homo sapiens.	XX	XX PP 10-DEC-2002; 2002WO-US039555.	XX	XX PP 10-DEC-2002; 2002WO-US039555.
XX US2003108915-A1.	PN	XX PR 10-DEC-2001; 2001US-0339739P.	XX	XX PR 10-DEC-2001; 2001US-0339739P.
XX PD 12-JUN-2003.	XX	XX PR 11-DEC-2001; 2001US-0339453P.	XX	XX PR 11-DEC-2001; 2001US-0339453P.
XX PF 20-AUG-2002; 2002US-00224624.	XX	XX PR 14-MAR-2002; 2002US-0365091P.	XX	XX PR 14-MAR-2002; 2002US-0365091P.
XX PR 20-OCT-2000; 2000US-0242160P.	XX	XX PR 14-MAR-2002; 2002US-0365384P.	XX	XX PR 14-MAR-2002; 2002US-0365384P.
PR 20-OCT-2001; 2001US-00051769.	PR	XX PR 12-APR-2002; 2002US-037381P.	PR	XX PR 12-APR-2002; 2002US-037381P.
XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.	PA	XX PR 22-APR-2002; 2002US-0379615P.	PR	XX PR 22-APR-2002; 2002US-0379615P.
XX Mckinnon RD;	PT	XX PR 24-APR-2002; 2002US-0376045P.	PR	XX PR 24-APR-2002; 2002US-0376045P.
XX WPI: 2003-810848/76.	XX	XX PA (HYSEQ INC.	PA	XX PA (HYSEQ INC.
XX DR; 2003-810848/76.	XX	XX PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;	XX	XX PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
XX Novel glioblastoma multiforme associated protein Gliten useful for	PT	XX PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;	PT	XX PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
treating glioblastoma multiforme and diagnosing brain cancer.	PT	XX PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;	PT	XX PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX Example 2; Fig 1; Opp; English.	PS	XX DR WPI: 2003-569235/53.	DR	XX DR WPI: 2003-569235/53.
XX The present invention provides novel glioblastoma multiforme (GBM)	XX	XX DR-PSDB; ADE08700.	DR	XX DR-PSDB; ADE08700.
CC associated protein Gliten useful for treating glioblastoma multiforme and	CC	XX PT New polynucleotides, useful for expressing recombinant proteins for	PT	XX PT New polynucleotides, useful for expressing recombinant proteins for
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,	CC	XX PT analysis, characterization or therapeutic use, or as markers for tissues	PT	XX PT analysis, characterization or therapeutic use, or as markers for tissues
CC treatment and diagnosis. The invention is also useful in gene therapy.	CC	XX PT in which the corresponding protein is preferentially expressed.	PT	XX PT in which the corresponding protein is preferentially expressed.
CC The present sequence is human Gliten partial cDNA. The human Gliten gene	CC	XX PS Claim 1; SEQ ID NO 855; 1177PP; English.	PS	XX PS Claim 1; SEQ ID NO 855; 1177PP; English.
CC is located on chromosome 10. This partial cDNA is used in the	CC	XX CC The invention comprises the amino acid and coding sequences of novel	CC	XX CC The invention comprises the amino acid and coding sequences of novel
CC exemplification of the invention.	CC	XX CC proteins. The DNA and protein sequences of the invention are useful as:	CC	XX CC proteins. The DNA and protein sequences of the invention are useful as:
XX Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;	SQ	XX CC markers for tissues in which the corresponding protein is preferentially	CC	XX CC markers for tissues in which the corresponding protein is preferentially
XX Query Match 100.0%; Score 261; DB 9; Length 960;	Db	XX CC expressed; as molecular weight markers on gels; as chromosome markers or	CC	XX CC expressed; as molecular weight markers on gels; as chromosome markers or
XX Best Local Similarity 100.0%; Pred. No. 1.2e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX CC tags; to identify chromosomes or to map related gene positions; and to	CC	XX CC tags; to identify chromosomes or to map related gene positions; and to
XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX CC compare with endogenous DNA sequences in patients to identify potential	CC	XX CC compare with endogenous DNA sequences in patients to identify potential
XX The present sequence is human Gliten partial cDNA. The present DNA sequence represents a gene of the	Qy	XX CC genetic disorders. The present DNA sequence represents a gene of the	CC	XX CC invention.
XX invention.	Qy	XX XX Sequence 1092 BP; 190 A; 377 C; 322 G; 203 T; 0 U; 0 Other;	XX	XX XX Sequence 1092 BP; 190 A; 377 C; 322 G; 203 T; 0 U; 0 Other;
XX Query Match 100.0%; Score 261; DB 9; Length 1092;	Db	XX XX Query Match 100.0%; Score 261; DB 9; Length 1092;	Qy	XX XX Query Match 100.0%; Score 261; DB 9; Length 1092;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX The present sequence is human Gliten partial cDNA. The present DNA sequence represents a gene of the	Qy	XX XX The present sequence is human Gliten partial cDNA. The present DNA sequence represents a gene of the	Qy	XX XX The present sequence is human Gliten partial cDNA. The present DNA sequence represents a gene of the
XX invention.	Qy	XX XX invention.	Db	XX XX invention.
XX Sequence 1 GATCAAGGTGAGTTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Sequence 1 GATCAAGGTGAGTTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Sequence 1 GATCAAGGTGAGTTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 396 GATCAAGGTGAGTTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Qy 61 GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 456 GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 121 CGTCGGCGAGTCGTCGCTCGCCACGGGAGCCGATGCCCTGACGGGGGCTT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 516 CGTCGGCGCTGGCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 120; DB 9; Length 120;	Db	XX XX Length 120; DB 9; Length 120;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 576 GCCCCGCGCTGCCATCGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 241 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 636 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 485; DB 9; Length 485;	Db	XX XX Length 485; DB 9; Length 485;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Score 61; GCTGAGCTCTGGAGCTGGTCTGGAGACGAGCTGGCTCATAGACCATTGGCT
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Db	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Length 545; DB 9; Length 545;	Db	XX XX Length 545; DB 9; Length 545;
Qy 181 GCCCCGCGCTGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	XX XX Best Local Similarity 100.0%; Pred. No. 1.1e-111; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 655 CCACCTGGCCATCGAGCTGGAACTGGCTTGAGGGAGCTGGCTCATAGACCATTGGCT	Db	XX XX Length 545; DB 9; Length 545;	Db	XX XX

RESULT 7  
 ADB62530  
 ID ADB62530 standard; cDNA; 2639 BP.  
 XX  
 AC ADB62530;  
 XX  
 DT 04-DEC-2003 (first entry)  
 DE Human cDNA encoding clone HHDPc20088160.  
 XX  
 KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
 tissue regeneration; cell regeneration; membrane protein;  
 signal transduction-related protein; transcription-related protein;  
 osteoporosis; neurological disease; cancer; tumour.  
 XX  
 OS Homo sapiens.  
 FH Key  
 FT CDS  
 FT /\*tag= a  
 FT /partial  
 FT /product= "Clone HHDPc20088160 protein"  
 XX  
 EP1308459-A2.  
 XX  
 PD 07-MAY-2003.  
 XX  
 PP 28-MAR-2002; 2002EP-00007401.  
 XX  
 PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350978.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PT Seki N, Yoshioka T, Otsuka M, Nagahari K, Matsuo Y;  
 XX  
 WPI: 2003-450961/43.  
 DR P-PSDB; ADB64500.  
 XX  
 PS 222pp; English.  
 XX  
 PT New polynucleotides and polypeptides, useful for developing a diagnostic  
 marker or medicines for regulation of their expression and activity, or  
 as targets of gene therapy.  
 XX  
 PS Claim 1; Page: 222pp; English.  
 XX  
 CC The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully modified nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or peptide  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The Oligonucleotide  
 CC is useful as a primer for synthesising the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,

CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
 CC data for this patent is not represented in the printed specification, but  
 CC is based on sequence information supplied by the European Patent Office.

AX Sequence 2639 BP; 657 A; 702 C; 719 G; 561 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 261; DB 9; Length 2639;  
 Best Local Similarity 100.0%; Pred. No. 1..1..11;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGACTTCAGGAGCTGCTGAGACCAAGACGCCGGCGCTGCTGAGGG 60  
 Db 579 GATCAAATGGATTGAGGTCAAGGAGGTGCTGAGCAAGAGGGGGGGCTGTGGGG 638

QY 61 GCTGAGCCTGCGGACCTGTTCTGGCGAGGTCGCCTCATCAAAGACCATCCGCT 120  
 Db 639 GCTGAGCCTGCGGACCTGTTCTGGCGAGGTCGCCTCATCAAAGACCATCCGCT 698

QY 121 CGTGCAGCGAGTGTGCTCGCCCATGGGAGGCCATGGCCTGAAGGGAGGGCT 180  
 Db 699 CGTGCAGCGAGTGTGCTCGCCCATGGGAGGCCATGGCCTGAAGGGAGGGCT 758

QY 181 GCCGGCCGCTGCCCCAGGGAGTGTGCTCGCCCATGGGAGGCCATGGCCTGAAGGGAGGGCT 240  
 Db 759 GCCGGCCGCTGCCCCAGGGAGTGTGCTCGCCCATGGGAGGCCATGGCCTGAAGGGAGGGCT 918

QY 241 CCACCTGGCCATCGACCTGGA 261  
 Db 819 CCACCTGGCCATCGACCTGGA 839

RESULT 8  
 ADD61564 standard; cDNA; 3832 BP.  
 XX  
 ID ADD61564 standard; cDNA; 3832 BP.  
 XX  
 AC ADD61564;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Human Gliten cDNA.  
 XX  
 Human glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;  
 XX  
 gene; glioblastoma; chromosome 10; gene; ss.  
 XX  
 Homo sapiens.  
 XX  
 FH Key  
 FT CDS  
 FT /\*tag= a  
 FT /product= "Human Gliten Protein"  
 FT /note= "The CDS is specifically claimed in claim 1"  
 FT primer\_bind  
 FT /\*tag= b  
 FT /bound\_moiety= "5' Primer #843"  
 FT 1011 .1032  
 FT primer\_bind  
 FT /\*tag= c  
 FT /bound\_moiety= "5' Primer #1405"  
 FT 1405 .1429  
 FT primer\_bind  
 FT /\*tag= d  
 FT /bound\_moiety= "5' Primer #2003"  
 FT 2083 .2108  
 FT primer\_bind  
 FT /\*tag= e  
 FT /bound\_moiety= "5' Primer #22355"  
 FT 3625 .3650  
 FT primer\_bind  
 FT /\*tag= f  
 FT /bound\_moiety= "3' Primer #3656"  
 XX  
 US2003108915-A1.  
 XX  
 12-JUN-2003.

PF 20-AUG-2002; 2002US-00224624.  
 XX PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00653661.  
 PR 20-OCT-2000; 2000US-00693125.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX PA (HYSE-) HYSEQ INC.  
 DR XX  
 PA (HYSE-) HYSEQ INC.  
 PT Tang YT, Liu C, Dumanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PT Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PT XX  
 PI XX  
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Goodrich R;  
 XX WPI: 2001-476283/51.  
 XX DR P-PSDB; AAM74475.  
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX PT  
 PS XX  
 PS Claim 1; Page 862-867; 6221pp; English.  
 XX CC The invention relates to polymucleotides (AAKS1456-AAKS3435) and the  
 CC encoded polypeptides (AAM78323-AAM0302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The invention or which may induce  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haemopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibition activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records ID NO 2110 (AAKS2581), 2111  
 CC (AAKS2582) and 3666 (AAM00020) are omitted from the relevant pages from the  
 CC sequence listing were missing at the time of publication.  
 XX SQ Sequence 3832 BP; 1090 A; 926 C; 957 G; 859 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 261; DB 9; Length 3832;  
 Best Local Similarity 100.0%; Pred. No. 1..e-111;  
 Matches 261; Conservative 0; Mismatches 0; Gaps 0;  
 Qy 1 GATCAAAGTGGAGTTCCAGGACTGTGCTGAGACCAAGACGCCAGTCAGAGGG 60  
 Db 543 GATCAAAGTGGAGTTCCAGGACTGTGCTGAGACCAAGACGCCAGTCAGAGGG 602  
 Qy 61 GTGAGGCTGGGAGCTGGCTGGCGAGAGGGTCCCTCATGAGACCATCGGCT 120  
 Db 603 GCTGAGGCTGGGAGCTGGTCCCTGGCGAGAGGGTCCCTCATGAGACCATCGGCT 662  
 Qy 121 CGTGGCCCASTGTGCCCTGGCCACGGGAGGCCATGGGAGGGCGCT 180  
 Db 663 CGTGCACGACAGTCGTGCCTGGCAACGGGAGGCCATGGGAGGGCGCT 722  
 Qy 181 GCCCGCGGCCCTCCCGAGGACTGTGGCTTCAGGGGGAGTACAAGGGGGCTT 240  
 Db 723 GCGCGCCCTACCCGAAGGTGGCTTCAGGGAGGGAGTACAAGGGGGCTT 782  
 Qy 241 CCACCTGGCCATGAGCTGGA 261  
 Db 783 CCACCTGGCCATGAGCTGGA 803  
 RESULT 9  
 AAK51608  
 ID AAK51608 standard; cDNA; 4450 BP.  
 AC AAK51608;  
 XX DT 06-NOV-2001 (first entry)  
 XX Human polynucleotide SEQ ID NO 153.  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haemopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157190-A2.  
 XX PD 09-AUG-2001.  
 XX PP 05-FEB-2001; 2001WO-US004098.  
 RESULT 10  
 ABA09174  
 ID ABA09174 standard; cDNA; 4470 BP.  
 AC ABA09174;  
 XX DT 11-JAN-2002 (first entry)

DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.

XX Human; cytokine; cell proliferation; cell differentiation; tissue growth; immunomodulator; activin;

KW haemopoiesis regulation; tissue growth; thrombolytic; oncogenesis;

KW inhibin; chemokines; chemokinesis; cancer; tumour; haematopoietic disorder;

KW proliferation; metatasis; lymphoid cell disorder; asthma; arthritis;

KW myeloid cell disorder; chronic inflammatory condition; arterial ischaemia;

KW atherosclerosis; coronary heart disease; arterial ischaemia;

KW bone disorder; osteoporosis; vascular growth disorder;

KW tissue regeneration; wound healing; infection; immune disorder;

KW cell culture; drug screening; gene therapy; antiinflammatory;

KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;

KW cytostatic; osteopathic; vasotrophic; cardiant; virucide; antibacterial;

KW antifungal; vulnerary; antiulcer; ss.

XX Homo sapiens.

XX WO200157188-A2.

XX PN 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US003800.

XX 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

XX PA (HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-457740/49.

DR P-PSDB; ABR11930.

XX PT Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis or cancer.

XX PS Claim 1; Page 813-814; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity;

CC haemopoiesis regulatory activity; tissue growth activity;

CC immunomodulatory activity; activin- or inhibin-related activities;

CC chemoatactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders.

CC Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness,

CC autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention

XX SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 4;

Best Local Similarity 100.0%; Pred. No. 1.e-111;

Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGAAGTTCAAGGAGCTGTGCATGACGCCAGCTGGAGGG 60

Db 366 GATCAAGTGAAGTTCAAGGAGCTGTGCATGACGCCAGCTGGAGGG 425

QY 61 GCTGAGCCCTGGGGACGGTTCCTGGGGAGCGGAGGGAGGGCT 120

Db 426 GCTGAGCCCTGGGGACGGTTCCTGGGGAGCGGAGGGAGGGCT 485

QY 121 CGTGCGCCAAGTGTGCCCTGGCACGGGAGGCCAGCCCTGAAGGGAGGGCT 180

Db 486 CGTGCGCCAAGTGTGCCCTGGCACGGGAGGCCAGCCCTGAAGGGAGGGCT 545

QY 181 GCGCGGCCCTGCCCGAGGACTGGCCTCGAGGCCAGGTGGAGTAACGGGGCTT 240

Db 546 GCGCGGCCCTGCCCGAGGACTGGCCTCGAGGCCAGGTGGAGTAACGGGGCTT 605

QY 241 CCACCTGGCATATGACCTGGA 261

Db 606 CCACCTGGCATATGACCTGGA 626

RESULT 11

AAK52592

ID AAK52592 standard; cDNA; 4470 BP.

XX AC AAK52592;

XX DT 06-NOV-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 2121.

XX KW Human; cytokine; cell proliferation; gene therapy;

XX KW vaccine; peptide therapy; stem cell growth factor; haemopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorder; arthritis; inflammation; ss.

XX OS Homo sapiens.

XX WO200157190-A2.

XX PD 09-AUG-2001.

XX PP 05-FEB-2001; 2001WO-US004098.

XX PR 03-FEB-2000; 2000US-00496314.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620125.

PR 01-SEP-2000; 2000US-00654336.

PR 15-SEP-2000; 2000US-0063361.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

XX (HYSEQ INC.

XX PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y; PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW; PI Xie AJ, Yang Y, Weijhrman T, Goodrich R;

AX DR WPI; 2001-476283/51.

AX DR P-PSDB; AAM79459.

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX Claim 1; Page 4496-4497; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 peptide therapy. The polypeptides have various cytokine-like activities,  
 e.g. stem cell growth factor activity, haemopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK5282) and 3666 (AAM80120) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication

XX Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 4; Length 4470;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-111;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATCAAGGTGAGTTCTGGAGCTTGACGAGCTGGTCAAGAACGAGCCGGCTGGAGGG 60  
 Db 366 GATCAAGGTGAGTTCTGGAGCTGGTCAAGAACGAGCCGGCTGGTCAAGAACGAGCCGG 425  
 Qy 61 GCTGAGSCCTGCCGGGACCTGTCCTGCCGAGCAGCTGGTCAAGAACGAGCCGGCTGGTCAAGAACGAGCCGGCT 120  
 Db 426 GCTGAGSCCTGCCGGGACCTGTCCTGCCGAGCAGCTGGTCAAGAACGAGCCGGCT 485  
 Qy 121 CGTGCAGCAGCTCGTGCAGCTCGCCACCGGGAGCCGATGCCCTGAAGGGAGCCGCT 180  
 Db 486 CGTGCAGCAGCTCGTGCAGCTCGCCACCGGGAGCCGATGCCCTGAAGGGAGCCGCT 545  
 Qy 181 GCCCCGCCCTGCCCGAGGAGCTGCGCTTGAGGGAGCTGGAGTCAACGGGGCTT 240  
 Db 546 GCCCCGCCCTGCCCGAGGAGCTGCGCTTGAGGGAGCTGGAGTCAACGGGGCTT 605  
 Qy 241 CCACCTGGCATTCGAGCTGGA 261  
 Db 606 CCACCTGGCATTCGAGCTGGA 626

RESULT 12  
 ADB09891 standard; DNA; 4470 BP.  
 XX AC ADE09891;  
 XX DT 29-JAN-2004 (first entry)

XX Novel DNA-related contig nucleotide sequence #613.  
 XX novel gene; novel protein; tissue marker; molecular weight marker;  
 KW chromosome marker; genetic disorder; contig; ds.  
 OS Unidentified.  
 PN WO2003054152-A2.  
 XX AC 03-JUL-2003.  
 XX PD 10-DEC-2002; 2002WO-US039555.

XX PR 10-DEC-2001; 2001US-0339739P.  
 PR 11-DEC-2001; 2001US-0339453P.  
 PR 14-MAR-2002; 2002US-0365091P.  
 PR 14-MAR-2002; 2002US-0365384P.  
 PR 12-APR-2002; 2002US-0372381P.  
 PR 12-APR-2002; 2002US-0372615P.

PR 22-APR-2002; 2002US-00128558.  
 PR 24-APR-2002; 2002US-0376045P.  
 XX PA (HYSE ) HYSEQ INC.  
 XX PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
 XX DR WPI; 2003-569235/53.  
 PT New polynucleotides, useful for expressing recombinant proteins for  
 PT analysis, characterization or therapeutic use, or as markers for tissues  
 PT in which the corresponding protein is preferentially expressed.  
 XX Disclosure; SEQ ID NO 2435; 117pp; English.  
 XX PS The invention comprises the amino acid and coding sequences of novel  
 CC proteins. The DNA and protein sequences of the invention are useful as:  
 CC markers for tissues in which the corresponding protein is preferentially  
 CC expressed, as molecular weight markers on gels; as chromosome markers or  
 CC tags; to identify chromosomes or to map gene positions; and to  
 CC compare with endogenous DNA sequences in patients to identify potential  
 CC genetic disorders. The present DNA sequence was used in the  
 CC exemplification of the invention.  
 XX SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 261; DB 9; Length 4470;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-111;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GATCAAGGTGAGTTCTGGAGCTGGTCAAGAACGAGCCGGCTGGTCAAGAACGAGCCGGCTGGTCAAGAACGAGCCGG 60  
 Db 366 GATCAAGGTGAGTTCTGGAGCTGGTCAAGAACGAGCCGGCTGGTCAAGAACGAGCCGGCTGGTCAAGAACGAGCCGG 425  
 Qy 61 GCTGAGSCCTGCCGGGACCTGTCCTGCCGAGCAGCTGGTCAAGAACGAGCCGGCTGGTCAAGAACGAGCCGGCT 120  
 Db 426 GCTGAGSCCTGCCGGGACCTGTCCTGCCGAGCAGCTGGTCAAGAACGAGCCGGCTGGTCAAGAACGAGCCGGCT 485  
 Qy 121 CGTGCAGCAGCTCGTGCAGCTCGCCACCGGGAGCCGATGCCCTGAAGGGAGCCGCT 180  
 Db 486 CGTGCAGCAGCTCGTGCAGCTCGCCACCGGGAGCCGATGCCCTGAAGGGAGCCGCT 545  
 Qy 181 GCCCCGCCCTGCCCGAGGAGCTGCGCTTGAGGGAGCTGGAGTCAACGGGGCTT 240  
 Db 546 GCCCCGCCCTGCCCGAGGAGCTGCGCTTGAGGGAGCTGGAGTCAACGGGGCTT 605  
 Qy 241 CCACCTGGCATTCGAGCTGGA 261  
 Db 606 CCACCTGGCATTCGAGCTGGA 626

RESULT 13  
 ADD49052  
 ID ADD49052 standard; DNA; 4801 BP.  
 XX AC ADD49052;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Human NOV6a coding sequence, SEQ ID 25.  
 XX KW Antidiabetic; anorectic; cardiotonic; hypotensive; antiarteriosclerotic;  
 KW virucide; antibacterial; fungicide; protozoic; nootropic;  
 KW neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;  
 KW antiarthritic; antiinflammatory; dermatological; antiasthmatic;  
 KW antilipemic; gene therapy; NOV protein; metabolic disorder; diabetes;  
 KW obesity; viral infection; bacterial infection; fungal infection;  
 KW helminthic infection; protozoal infection; anoxia; cancer;  
 KW cardiovascular disease; hypertension; atherosclerosis;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW epilepsy; immune disorder; osteoarthritis; haemato poetic disorder;

KW inflammatory skin disorder; asthma; dyslipidemia; human; gene; ds  
 XX OS Homo sapiens.  
 XX PN WO2003060149-A2.  
 XX PR 24-JUL-2003 .  
 XX PR 06-JAN-2003 ; 2003WO-US0000252 .  
 XX PR 04-JAN-2002 ; 2002US-0345222P .  
 XX PR 14-JAN-2002 ; 2002US-034693P .  
 XX PR 16-JAN-2002 ; 2002US-034918P .  
 XX PR 17-JAN-2002 ; 2002US-034973P .  
 XX PR 18-JAN-2002 ; 2002US-035066P .  
 XX PR 24-JAN-2002 ; 2002US-035197P .  
 XX PR 28-MAY-2002 ; 2002US-038375P .  
 XX PR 05-JUN-2002 ; 2002US-035963P .  
 XX PR 11-JUN-2002 ; 2002US-038783P .  
 XX PR 17-JUL-2002 ; 2002US-039640P .  
 XX PR 30-SEP-2002 ; 2002US-041511P .  
 XX PR 03-JAN-2003 ; 2003US-03336603 .  
 XX PA (CTRA-) CURAGEN CORP .  
 XX PPI Grossé WM, Alsobrook JP, Anderson DW, Burgess CE, Edinger SR;  
 PPI Ellerman K, Furtak K, Gangolli EA, Gerlach VL, Gilbert JA;  
 PPI Gunther E, Gorman L, Guo X, Ji W, Li L, Miller CB, Padigaru  
 PPI Patterson M, Rastelli L, Macdougall JR, Mishra VS, Smithson G,  
 PPI Spyres KA, Stone DJ, Sheroy SC, Taupier RJ, Vernet CAM,  
 PPI Malayankar UM, Millet I, Rekuda R,  
 PPI XX  
 DBP: 2003-5877299/EE

QY	121	CCTGGGCCCCATGCCCTCGACCACTGGGAGCCGATGGCCCTGAAGGGGGCT 180
Db	663	CGTGGGCCATGCCCTCGACCACTGGGAGCCGATGGCCCTGAAGGGGGCT 722
QY	181	GCCCCCCTGCCTGCGAGGTGGCTTICGAGGGAGGTGAGGTACAACGGGGCT 240
Db	723	GCCCCCCTGCCCTGCCAGGAGCTGCCCTGAGGTGAGGTACAACGGGGCT 782
QY	241	CCACTGCCCATTGCGTGA 261
Db	783	CCACTGCCCATTGCGTGA 803
<b>RESULT 14</b>		
	ABK87764	standard; cDNA; 3B47 BP.
XX	ID	ABK87764 standard; cDNA; 3B47 BP.
XX	AC	ABK87764;
XX	DT	24-SEP-2002 (first entry)
XX	DE	Human cDNA encoding glioma antigen KU-GB-5.
XX	KW	Human; ss; gene; glioma antigen; KU-GB-5; glioma; cytostatic; cancer; immunostimulant; immunosuppressive; peptide therapy; gene therapy; malignant brain tumour.
XX	KW	
XX	KW	
OS		Homo sapiens.
XX		Location/Qualifiers
FH	Key	211 .2883
FT	CDS	/*tag= a
FT		/product= "KU-GB-5"
FT		
XX	PN	WO200255695-A1.
XX	PD	18-JUL-2002.
XX	PR	30-NOV-2001; 2001WO-JP010505.
XX	PR	09-JAN-2001; 2001JP-000001965.
XX	PA	(UYKE-) UNIV KEIO.
XX	PI	Toda M, Kawakami Y, Kawase T, Iizuka Y;
XX	DR	WPI; 2002-538141/57.
DR	P-PDB; AAU99614.	
XX	PT	New human glioma antigen for diagnosing and treating glioma, human malignant brain tumor and other cancers, and for studying onset of glioma.
XX	PS	Claim 12; Page 94-100; 109pp; Japanese.
XX	CC	The invention relates to preparing glioma antigen and/or glioma antigen gene comprises e.g. extracting and isolating total RNA from a glioma cell line then synthesising cDNA for constructing a lambda phage cDNA library and reacting the library with the serum for reaction and detecting positive clones reacting with the antibody in the serum by using a labeled anti-immunoglobulin (IgG antibody). Also included are diagnostic reagents for detecting glioma containing 1 or more kinds of the whole or partial glioma antigens thus prepared, and/or an antibody binding with the glioma antigens and/or parts of them, probes for detecting or diagnosis of glioma containing the whole or partial antisense chains of DNAs or RNAs encoding the prepared whole or partial glioma antigens, DNAs (or cDNAs) encoding the glioma antigens (AAU99609-AAU99614) and the protein sequences for the antigens (peptides) for peptides derived from them, a DNA hybridisable with the DNA in under stringent conditions and encoding a protein with immune induction activity, a fusion protein or peptide formed by bonding the protein or its partial peptide with a marker peptide tag, an antibody for the protein or its partial peptide tag, a host cell containing an expression system for expressing the

protein or peptide, an non-human animal which has deletion of the gene function encoding the protein or its partial peptide on the chromosome, or has overexpression of the protein or its partial peptide and screening substances promoting or inhibiting immune induction activity by using the protein or its partial peptide, the test substances and T cells, and measuring and evaluating immune induction activity in T cells. The antigen and DNA encoding it, are applicable in diagnosis and treatment of (e.g. by peptide or gene therapy) glioma, human malignant brain tumour and other cancers, and for studying onset of glioma. The present sequence encodes the glioma antigen KU-0B-5

```
XX Sequence 3847 BP; 1197 A; 744 C; 819 G; 1086 T; 0 U; 1 Other;
SQ Query Match 18.0%; Score 47; DB 6; Length 3847;
Best Local Similarity 100.0%; Pred. No. 3 7e-12;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 215 GCGGAGGTGGAGTACAACGGGGCTTCACCTGGCCATCGACTGGA 261
Db 1 GCGGAGGTGGAGTACAACGGGGCTTCACCTGGCCATCGACTGGA 47
```

```
RESULT 15
ABX95681
ID ABX95681 standard; cDNA; 263 BP.
XX AC ABX95681;
XX DT 30-JUN-2003 (First entry)
XX DE Rat Gliten expressed sequence tag.
XX KW Rat; ss: EST; Glioblastoma multiforme; brain cancer; Gliten; GMB;
expressed sequence tag.
XX OS Rattus Norvegicus.
XX PN US2003044811-A1.
XX PD 06-MAR-2003.
XX PF 20-OCT-2001; 2001US-00051769.
XX PR 20-OCT-2000; 2000US-0242160P.
XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX PT McKinnon RD;
DR WPI; 2003-418995/39.
```

```
XX PT New isolated nucleic acid representing a gene product associated with
Glioblastoma Multiforme, designated as Gliten, useful as a molecular
marker for tumor cell identification and classification.
XX PS Example 2; Page 4; 11pp; English.
XX PT The invention relates to an isolated nucleic acid representing a gene
product associated with Glioblastoma Multiforme, designated as Gliten,
comprising the human EST (expressed sequence tag) appearing as ABX95678,
or a sequence that hybridises under stringent conditions to the EST, or
its complement. Also included are a probe for use in identifying
patient at risk for progression into the malignant phenotype, comprising
the nucleic acids detailed above, detecting whether a patient is at risk
for progression into Glioblastoma Multiforme (comprising: (a) providing a
sample from a patient; (b) adding the probe to the sample or performing
PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
analysing levels of mRNA bound with the probe; and (d) treating a control
sample to assess the level of mRNA in a control sample, where the
presence of increased levels of mRNA expression in the sample in an
amount higher than the control sample indicates risk for progression into
Glioblastoma Multiforme), and a kit for use in detecting whether a
patient is at risk for progression into Glioblastoma Multiforme
```

```
CC comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is a Rat Gliten expressed sequence tag
XX Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;
SQ Query Match 16.5%; Score 43; DB 7; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.1e-10;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GATCAGGGTGGAGTTCGAGGAGCTGCTGCAAGACGCC 43
Db 1 GATCAAGTGGAGTTCGAGGCTGCTGCAAGACGCC 43
```

```
Search completed: July 18, 2004, 14:51:45
Job time : 250 secs
```

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

	Result No.	Score	Query Match Length	DB ID	Description
Run on:	July 18, 2004, 13:33:30 ; Search time 1612 Seconds (without alignments)				
	7017.700 Million cell updates/sec				
Title:	US-10-051-769-2				
Perfect score:	261				
Sequence:	1 gatccaagggtggacttcagg... . . . . . caoctggccatgcacrggaa 261				
Scoring table:	OLIGO NUC				
Gapext 60.0 , Gapext 60.0					
Searched:	3470272 seqs, 21671516995 residues				
Word size :	0				
Total number of hits satisfying chosen parameters:	6940544				
Minimum DB seq length: 0					
Maximum DB seq length: 2000000000					
Post-processing: Listing first 45 summaries					
Database :	GenEmbl:*				
1: gb_ba:*					
2: gb_htg*:*					
3: gb_in:*					
4: gb_on:*					
5: gb_ov:*					
6: gb_pa:*					
7: gb_ph:*					
8: gb_pl:*					
9: gb_pr:*					
10: gb_xo:*					
11: gb_sts:*					
12: gb_sy:*					
13: gb_un:*					
14: gb_vr:*					
15: em_ba:*					
16: em_fun:*					
17: em_hum:*					
18: em_htg_other:*					
19: em_ir:*					
20: em_on:*					
21: em_oi:*					
22: em_cv:*					
23: em_pa:*					
24: em_ph:*					
25: em_pl:*					
26: em_rc:*					
27: em_sts:*					
28: em_ur:*					
29: em_vl:*					
30: em_htg_hum:*					
31: em_htg_inv:*					
32: em_htg_other:*					
33: em_htg_mus:*					
34: em_htg_pln:*					
35: em_htg_rod:*					
36: em_htg_man:*					
37: em_htg_vrl:*					
38: em_sy:*					
39: em_htgo_hum:*					
40: em_htgo_mus:*					
41: em_htgo_other:*					

## ALIGNMENTS

RESULT 1				
BC012186				
LOCUS	Homo sapiens, clone	2623 bp	mRNA	linear
DEFINITION		IMAGE:454853,	mRNA	PRI 06-AUG-2001
ACCESSION	BC012186			
VERSION	BC012186.1	GI:15082555		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
Eukaryota; Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Butheria; Primates; Catarrini; Hominidae; Homo.				
1 (bases 1 to 2623)				
Strausberg, R.				
Direct Submission				
Submitted (02-AUG-2001) National Institutes of Health, Mammalian				

Pred. No. is the number of results predicted by chance to have a

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2390, USA

NHG-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

info@bcgsc.ca

Steven Chan, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reka Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saiedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Teai, Natasha van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marras.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: ITRAL Plate: 29 Row: h Column: 17  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomeScan gene prediction.

Location/Qualifiers

1..2623

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="Taxon:9606"

/clone="IMAGE:4564953"

/tissue="Type=Kidney, renal cell adenocarcinoma"

/clone\_Lib="NIH MGC 14"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

ORIGIN

Query Match 100.0% Score 261; DB 9; Length 2623;  
Best Local Similarity 100.0% Pred. No. 1e-120;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 543 GATCAAGTGGAGTTGAGCTGAGAACAGAACGCCGGCGCTGTGAGGG 60

Qy 61 GCTGAGCCCTGGGGAGCTGCTGGGAGACGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 602

Db 603 GCTGAGCCCTGGGGAGCTGCTGGGAGACGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 662

Qy 121 CTTGGGGCAGTCGCGCCCTGGGGAGCCGGAGCCGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 180

Db 663 CTTGGGGCAGTCGCGCCCTGGGGAGCCGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 722

Qy 181 GCCCGCCGCTGCCCCGAGAGCTGGCCCTTGAGGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 240

Db 722 GCGCGCGCTGCCCCGAGAGCTGGCCCTTGAGGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 782

Qy 241 CCACCTGGCCATCGACGGTGA 261

Db 783 CCACCTGGCCATCGACGGTGA 803

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

I Isogai,T., Sugiyama,T., Otuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshihiko,T., Masuho,Y.

Title Full-length cDNA sequences Patent: EP 130845-A 684 07-MAY-2003; Helix Research Institute (JP) ; Research Association for Biotechnology (JP)

Features Location/Qualifiers source 1..2639 /organism="Homo sapiens" /mol\_type="mRNA" /db\_xref="taxon:9606"

Origin Query Match 100.0%; Score 261; DB 6; Length 2639; Best Local Similarity 100.0%; Pred. No. 1e-120; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGTGGAGTTGAGCTGCTGGGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 60 Db 579 GATCAAGTGGAGCTGTGAGGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 638

Qy 61 GCTGAGCCCTGGGGAGCTGTGAGGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 120 Db 639 GCTGAGCCCTGGGGAGCTGTGAGGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 698

Qy 121 CGTGGCCCGAGTCGCGCCCTGGGGAGCTGTGAGGGAGCTGTGAGGGAGCTGTGAGGG 180 Db 699 CGTGGCCCGAGTCGCGCCCTGGGGAGCTGTGAGGGAGCTGTGAGGGAGCTGTGAGGG 758

Qy 181 GCCCGCCGCTGCCCCGAGAGCTGGCCCTTGAGGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 240 Db 759 GCGCGCCGCTGCCCCGAGAGCTGGCCCTTGAGGGAGCTGTGAGAACAGAACGCCGGCGCTGTGAGGG 818

Qy 241 CCACCTGGCCATCGACGGTGA 261 Db 819 CCACCTGGCCATCGACGGTGA 839

Result 3 AK091746 Locus AK091746 2639 bp mRNA linear PRI 15-JUL-2002

Definition Homo sapiens cDNA FLJ34427 fis, clone HDSPC200816.

Version AK091746 1 GI:21750192

Keywords oligo capping; fis (full insert sequence). Source Homo sapiens (human)

Organism Homo sapiens

Authors Sugiyama,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,A., Wakamatsu,A., Ishii,S., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Matsumoto,Y., Nagai,K., and Isogai,T.

Title Unpublished

Journal 2 (bases 1 to 2639)

Authors Isogai,T. and Yamamoto,J.

Title Direct Submission

Journal Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

Comment (E-mail:)genomicsbri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library

Result 2 AX747159

Locus AX747159

Definition Sequence 684 from Patent EP1308459.

Accession AX747159

Version AX747159.1 GI:32131547

Keywords Homo sapiens (human)

Organism Homo sapiens

TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	22388257		
PUBMED	12477932		
REPERE	2 (bases 1 to 3856)		
AUTHORS	Strausberg, R.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.ncbi.nlm.nih.gov">http://mgc.ncbi.nlm.nih.gov</a>		
COMMENT	Contact: MGC help desk Email: cgaps-r@mail.nih.gov		
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL) DNA Sequencing by: Institute for Systems Biology contact: <a href="http://www.systemsbiology.org">http://www.systemsbiology.org</a>		
	Anup Madan, Jessica Fahey, Brian Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting		
FEATURES	Location/Qualifiers	source	
	1. .3856	/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="MGC:27107 IMAGE:4837939"	
		/tissue_type="testis"	
		/clone_id="NIH_MGC_97"	
		/lab_host="DH10B"	
		note="vector: pBluescript"	
gene	1. .3856	/gene="LOC118987"	
CDS	203_.3667	/db_xref="LocusID:118987"	
		/codon_start=1	
		/product="hypothetical protein LOC118987"	
		/protein_id="AAH232375_1"	
		/db_xref="GI:22387224"	
		/translation="MGIIIMIILSAVGLSPFLTLAQFELLYRQPEPDEAAARGESFRIVKPVGGLLRKYLGGRDKEBPGAAPEGGATPTAEPETPAPPTRTCYFNUATLFLFREIRDATLTRWTKKQYKEFEEULLQTGAGRLLGSLURDFFUCLGTTVPIKTLFLVLPDUDNCGPDCPGEALACPCAPAEFLAAGCCTGGCCTTGTGAGGTCGKSYFLVPSRVVYGLRLUVFTRYFTHWFPSVBDPLDFEVRSQFEGRPMPQQTISITVNLKFLIKRKHTLPNTYKIRKFPPYOTLOGEDEBRHIIQWALTEGRKLWLLCSRLIICGSSYDRBANQKHTLELSSTWEEKRQSASIKTYVEIQLGNLQSQVYVLLVOSTGCTYQAHVLTETVANSPRAIAQDQEGLIAIGGKVKTSTLQVLKIICQADGRLVYTPRGVOSNOGAVIDQNFSCOLEMFLSSSCSGYEEAATLVDTESRESDLPDSEFDASDVRAQNEFKDAGLSHSPKRFTLISKPLQGSPVNRKQHSLPDKTQGDKNQPKPDPKQDQKQPKPLKTSCTTDQAQVRKPTQGSAFKPVPPQRAQVKLPSAAPNQBPDVIVEKPKVVPPLVDKSAEKQAKNVDADAQPAKPOFLAKQEVADYTSBTSCPTKDSDDDRCTWESESSLVLYRKQKHTTRASCLIEDAICHYTNALMRDPKQDQVTPRQVLSLQKJLGEVSKLXEDVAIGLQATSNTEYLSKLRLAAPSRAIVTALRNMSQDQFCKYGDITIHKYLKBLGEDDHVVYVNECEKEPHIUYEVSVLPREEQFYGMQGTMENKSFQDTEFQNPTWCYDCKKKWTAQASQCMFCAVCHKRCQKCAETTSGATDRDITLNRLEGDETLILLGIPRVDABAEASKVNKITGTLTRELIINTSRSLNROVERLTSRTEPQDTEPVSPTKPTNTSDNEGSDDTEVCGPNSSPSKRGNSTGKLVKVRKEGLDLSVEFIAKEIGDLYLRCPTERIQLPENMDQNEQDQLENNNSLVREPEEFTDTRKXKSLAALSGERSLTQDTEEDIDNETSQLIDSQPFSSISQD	REPERE
misc_feature	Region: PSD-95, Did, and PSD-1546		
	note="Region: PSD-95, Did, and PSD-1546"		

ZO-1/2. Also called DHR (DLG homologous region) or GIGF (relatively well conserved tetrapeptide in these domains). Some PDZs have been shown to bind C-terminal polypeptides." /db\_xref="CDD;smart00228"

**misc\_feature**  
2723 2851 /note "DAG PE-bind; Region: Phorbol esters/diacylglycerol binding domain (CL domain). This domain is also known as the Protein kinase C conserved region 1 (CL) domain" /db\_xref="CDD;pfam0130"

**ORIGIN**

	Query Match	Score	Length	Source	FEATURES	ORIGIN
	100.0%	261	DB 9;			Query Match 100.0% Score 261; DB 9; Length 49052;
Best Local Similarity	100.0%	Pred. No. 1e-120;				Best Local Similarity 100.0%; Pred. No. 8.4e-121;
Matches	261;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GATCAGGTGGAGTTCGAGGCTGTGCGACCAAGACGCCGCGCTGCTGAGGG 60				QY 1 GATCAGGTGGAGTTCGAGGAGGTCTGAGACCAAAGACGCCGCGCTGCTGAGGG 60
Db	568	GATCAGGTGGAGTTCGAGGCTGTGCGACCAAGACGCCGCGCTGCTGAGGG 627				Db 29767 GATCAGGTGGAGTTCGAGGAGGTCTGAGACCAAAGACGCCGCGCTGCTGAGGG 29768
QY	61	GCTGAGCCCTGGGGAGCTGGTGTCTGGGGAGACGCGTGCCTTCATCAAACCATTGGCT 120				QY 61 GCTGAGCCCTGGGGAGCTGGTGTCTGGGGAGACGCGTGCCTTCATCAAACCATTGGCT 120
Db	628	GCTGAGCCCTGGGGAGCTGGTGTCTGGGGAGACGCGTGCCTTCATCAAACCATTGGCT 687				Db 29707 GCTGAGCCCTGGGGAGCTGGTGTCTGGGGAGACGCGTGCCTTCATCAAACCATTGGCT 29648
QY	121	CCTGGCGCCAGTCGAGCCCTGGCCACCGGGAGCCGCGCTGAGGGAGGGCGCT 180				QY 121 CCTGGCGCCAGTCGAGCCCTGGCCACCGGGAGCCGCGCTGAGGGAGGGCGCT 180
Db	688	CCTGGCGCCAGTCGAGCCCTGGCCACCGGGAGCCGCGCTGAGGGAGGGCGCT 747				Db 29647 CCTGGCGCCAGTCGAGCCCTGGCCACCGGGAGCCGCGCTGAGGGAGGGCGCT 29588
QY	181	GCCCCGCCCTGGCCCGAGGACTGGCCCTTGAGGGAGTGGAGTAAACGGGGCTT 240				QY 181 GCCCCGCCCTGGCCCGAGGACTGGCCCTTGAGGGAGTGGAGTAAACGGGGCTT 240
Db	748	GCCCCGCCCTGGCCCGAGGACTGGCCCTTGAGGGAGTGGAGTAAACGGGGCTT 807				Db 29587 GCCCCGCCCTGGCCCGAGGACTGGCCCTTGAGGGAGTGGAGTAAACGGGGCTT 29587
QY	241	CCACCTGGCATCGAGTGGAA 261				QY 241 CCACCTGGCATCGAGTGGAA 261
Db	808	CCACCTGGCATCGAGTGGAA 828				Db 29527 CCACCTGGCATCGAGTGGAA 29507
<b>RESULT 5</b>						
<b>AL359836/c</b>	<b>AL359836</b>	<b>49052 bp</b>	<b>DNA</b>	<b>PRI 21-DEC-2001</b>	<b>LOCUS</b>	<b>AC005887</b>
<b>DEFINITION</b>	<b>Human DNA sequence from clone RP11-389E6</b>	<b>on chromosome 10, complete sequence.</b>			<b>DEFINITION</b>	<b>citb113_i_12, complete sequence.</b>
<b>ACCESSION</b>	<b>AL359836</b>				<b>ACCESSION</b>	<b>AC005887</b>
<b>VERSION</b>	<b>AL359836.16</b>	<b>GI:17977720</b>			<b>VERSION</b>	<b>GI:6249675</b>
<b>KEYWORDS</b>					<b>KEYWORDS</b>	<b>HTG.</b>
<b>SOURCE</b>	<b>Homo sapiens (human)</b>				<b>SOURCE</b>	<b>Homo sapiens (human)</b>
<b>ORGANISM</b>	<b>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</b>				<b>ORGANISM</b>	<b>Homo sapiens</b>
<b>REFERENCE</b>	<b>1 (bases 1 to 49052)</b>				<b>REFERENCE</b>	<b>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</b>
<b>AUTHORS</b>	<b>Smith, M.</b>				<b>AUTHORS</b>	<b>Smith, D.R.</b>
<b>TITLE</b>	<b>Direct Submission</b>				<b>TITLE</b>	<b>Unpublished</b>
<b>JOURNAL</b>	<b>Submitted (21-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail inquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk</b>				<b>JOURNAL</b>	<b>Journal of Human Genetics</b>
<b>COMMENT</b>	<b>On Dec 23, 2001 this sequence version replaced Gi:17384082. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations corresponding to the overlapping clone name. Note that the variation annotation may not be found in the sequence submission with only a small overlap as described above.</b>				<b>COMMENT</b>	<b>2 (bases 1 to 120578)</b>
<b>ACCESSION</b>	<b>AL359836</b>				<b>ACCESSION</b>	<b>Smith, D.R.</b>
<b>VERSION</b>	<b>AL359836.16</b>				<b>VERSION</b>	<b>Direct Submission</b>
<b>SOURCE</b>	<b>Homo sapiens</b>				<b>SOURCE</b>	<b>Submitted (29-OCT-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154 , USA</b>
<b>ORGANISM</b>	<b>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</b>				<b>ORGANISM</b>	<b>3 (bases 1 to 120578)</b>
<b>REFERENCE</b>	<b>1 (bases 1 to 120578)</b>				<b>REFERENCE</b>	<b>Smith, D.R.</b>
<b>AUTHORS</b>					<b>AUTHORS</b>	<b>Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154 , USA</b>
<b>TITLE</b>					<b>TITLE</b>	<b>Direct Submission</b>
<b>JOURNAL</b>					<b>JOURNAL</b>	<b>Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154 , USA</b>
<b>COMMENT</b>					<b>COMMENT</b>	<b>3 (bases 1 to 120578)</b>
<b>ACCESSION</b>					<b>ACCESSION</b>	<b>Smith, D.R.</b>
<b>VERSION</b>					<b>VERSION</b>	<b>Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154 , USA</b>
<b>SOURCE</b>					<b>SOURCE</b>	<b>Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154 , USA</b>
<b>ORGANISM</b>					<b>ORGANISM</b>	<b>3 (bases 1 to 120578)</b>
<b>REFERENCE</b>					<b>REFERENCE</b>	<b>Smith, D.R.</b>
<b>AUTHORS</b>					<b>AUTHORS</b>	<b>Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154 , USA</b>
<b>TITLE</b>					<b>TITLE</b>	<b>Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154 , USA</b>
<b>JOURNAL</b>					<b>JOURNAL</b>	<b>Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154 , USA</b>

was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10 RP11-389E6 is from the library RPCI-11.2 constructed by the group of Pierie de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBAcE3.6

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

IMPORT: This sequence is not the entire insert of clone RP11-389E

REFERENCE	A61K45/00, A61K48/00, A61P35/00, C07K14/47, C07K16/18, C07K19/00,	PC	PC
AUTHORS	Smith,D.R.	C12Q1/02,	C12Q1/68
TITLE	Direct Submission	Human glioma antigen and method of preparing the same	Key
JOURNAL	Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA	Location/Qualifiers	
REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA	Location/Qualifiers	
AUTHORS	Smith,D.R.	Location/Qualifiers	
TITLE	Direct Submission	1. .3847	
JOURNAL	Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA	/organism="Homo sapiens"	
REMARK	Vector Sequence Clipped	/mol type="genomic DNA"	
COMMENT	On Nov 5, 1999 this sequence version replaced gi:4314331.	/db_xref="#taxon:9606"	
FEATURES	source	1. /map="10q25"	
ORIGIN	1. /clone_lib="CIT987SK-1173112"		
Query Match	Score 261; DB 9%; Length 120578;	Query Match	18.0%; Score 47; DB 6; Length 3847;
Best Local Similarity	100.0%; Pred. No. 7.8e-121;	Best Local Similarity	100.0%; Pred. No. 1e-12;
Matches	261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Matches	47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 GATCAAGGTGAGATTGGAGGTGCTGA GACCAAGGGCGCCCTGTGGAGGG 60	AC108407	AC108407 bp DNA linear HTG 27-JAN-2002
Db	53976 GATCAAGGTGAGATTGGAGGTGCTGA GACCAAGGGCGCCCTGTGGAGGG 54035	LOCUS	Mus musculus clone RP24-422P10, LOW-PASS SEQUENCE SAMPLING.
Qy	61 GCTGAGCTGGGACGGTGTCTGGGGAGGGCTCATGAGCCATCCGCCT 120	DEFINITION	Mus musculus
Db	54036 GCTGAGCTGGGACGGTGTCTGGGGAGGGCTCATGAGCCATCCGCCT 54095	ACCESSION	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Qy	121 CGTGCGGCCAGGTGTGGCCCTGGCCAACGGGGAGCCGATGGCCATGAAAGGGAGGGCT 180	VERSION	Mammalia; Buthidae; Rodentia; Sciurognathi; Muridae; Mus.
Db	54096 CGTGCGGCCAGGTGTGGCCCTGGCCAACGGGGAGCCGATGGCCATGAAAGGGAGGGCT 54155	KEYWORDS	1. (bases 1 to 68196)
Qy	181 GCGCGCCCTGCCCGAGGAGGTGGCTTGGGGAGGGCTT 240	ORGANISM	1. (bases 1 to 68196)
Db	54156 GCGCGCCCTGCCCGAGGAGGTGGCTTGGGGAGGGCTT 54215	REFERENCE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Qy	241 CCACCTGGCCATGACGGCTGA 261	AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Db	54216 CCACCTGGCCATGACGGCTGA 54236	TITLE	Unpublished
RESULT	7	JOURNAL	2. (bases 1 to 68196)
BD169701	BD169701	REFERENCE	Birren,B., Linton,L., Nusbaum,C., Landier,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgatier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,D., Collymore,A., Cook,A., Cooke,P., DeAngelis,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,J., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., McDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menes,L., Mihowa,T., Mlenga,V., Murphy,T., Mayor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Rieback,M., Riley,R., Raymond,C., Retta,R., Rieback,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,N., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembik,L., Zimmer,A. and Zody,M.
DEFINITION	Human glioma antigen and method of preparing the same.	COMMENT	All repeats were identified using RepeatMasker:
ACCESSION	BD169701	TITLE	Submitted (27-JAN-1996-1997)
VERSION	BD169701.1	JOURNAL	http://ftp.genome.washington.edu/RM/RepeatMasker.html
KEYWORDS	GI:27875513	COMMENT	----- Genome Center
SOURCE	WO 02055635-A/6	TITLE	Center: Whitehead Institute/ MIT Center for Genome Research
ORGANISM	Homo sapiens (human)	JOURNAL	320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	1. (bases 1 to 3847)	COMMENT	All repeats were identified using RepeatMasker:
AUTHORS	Toda M., Kawakami,Y., Kawase,T. and Iizuka,Y.	TITLE	Submitted (27-JAN-2002)
TITLE	Human glioma antigen and method of preparing the same	JOURNAL	Whitehead Institute/ MIT Center for Genome Research
JOURNAL	Patent: WO 02055635-A 6-18-JUL-2002;	COMMENT	Center code: WIBR
COMMENT	YUKIHIKO IIZUKA	TITLE	Web site: http://www-seq.wi.mit.edu
OS	Homo sapiens (human)	JOURNAL	Contact: sequence_submissions@genome.wi.mit.edu
PD	WO 02055635-A/6	COMMENT	----- Project Information
PF	30-NOV-2001 WO 2001JP010505	TITLE	Center project name: Li8881
PR	09-JAN-2001 JP 011P 001965	JOURNAL	
PI	MASAHIRO TODA,YUTAKA KAWAKAMI,TAKESHI KAWASE,YUKIHIKO IIZUKA	COMMENT	
PC	C12N15/12,C12N5/10,A01K67/027,A61K31/711,A61K38/00,A61K39/00,A61K39/395,	TITLE	

Center clone name: 422\_P\_10

\* NOTE: This record contains 85 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Lowpass sequence sampling is useful for identifying clones that may be gene rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1 714: contig of 714 bp in length

1	715	814:	gap of 100 bp	
*	815	1520:	contig of 706 bp in length	
*	1521	1620:	gap of 100 bp	
*	1621	2323:	contig of 703 bp in length	
*	2324	2433:	gap of 100 bp	
*	2424	3134:	contig of 711 bp in length	
*	3135	3233:	gap of 100 bp	
*	3235	3937:	contig of 703 bp in length	
*	3938	4037:	gap of 100 bp	
*	4038	4779:	contig of 702 bp in length	
*	4740	4839:	gap of 100 bp	
*	4840	5545:	contig of 706 bp in length	
*	5546	5645:	gap of 100 bp	
*	5646	6352:	contig of 707 bp in length	
*	6353	6452:	gap of 100 bp	
*	6453	7172:	contig of 720 bp in length	
*	7173	7272:	gap of 100 bp	
*	7273	7996:	contig of 724 bp in length	
*	7997	8096:	gap of 100 bp	
*	8097	8814:	contig of 718 bp in length	
*	8815	8914:	gap of 100 bp	
*	8915	9629:	contig of 715 bp in length	
*	9630	9729:	gap of 100 bp	
*	9730	10417:	contig of 688 bp in length	
*	10418	10517:	gap of 100 bp	
*	10518	11222:	contig of 706 bp in length	
*	11224	11322:	gap of 100 bp	
*	11324	12014:	contig of 691 bp in length	
*	12015	12114:	gap of 100 bp	
*	12115	12816:	contig of 702 bp in length	
*	12817	12916:	gap of 100 bp	
*	12917	13616:	contig of 700 bp in length	
*	13716	13716:	gap of 100 bp	
*	13717	14422:	contig of 706 bp in length	
*	14423	14522:	gap of 100 bp	
*	14523	15234:	contig of 712 bp in length	
*	15235	15333:	gap of 100 bp	
*	15335	16043:	contig of 691 bp in length	
*	16044	16143:	gap of 100 bp	
*	16144	16859:	contig of 716 bp in length	
*	16860	16959:	gap of 100 bp	
*	16960	17669:	contig of 710 bp in length	
*	17670	17769:	gap of 100 bp	
*	17770	18460:	contig of 691 bp in length	
*	18461	18560:	gap of 100 bp	
*	18561	19269:	contig of 709 bp in length	
*	19270	19369:	gap of 100 bp	
*	19370	20070:	contig of 701 bp in length	
*	20071	20170:	gap of 100 bp	
*	20171	20874:	contig of 704 bp in length	
*	20875	20974:	gap of 100 bp	
*	20975	21684:	contig of 709 bp in length	
*	21685	21784:	gap of 100 bp	
*	21785	22491:	contig of 707 bp in length	
*	22492	22591:	gap of 100 bp	
*	22592	23299:	contig of 708 bp in length	
*	23399:	gap of 100 bp		
*	23400	24086:	contig of 687 bp in length	
*	24087	24186:	gap of 100 bp	

*	24187	24901:	contig of 715 bp in length	
*	24902	25001:	gap of 100 bp	
*	25002	25714:	contig of 713 bp in length	
*	25715	25514:	gap of 100 bp	
*	25815	26528:	contig of 714 bp in length	
*	26529	26628:	gap of 100 bp	
*	26629	27338:	contig of 730 bp in length	
*	27359	28559:	gap of 100 bp	
*	28559	28559:	contig of 701 bp in length	
*	28666:	28666:	contig of 707 bp in length	
*	28867	29065:	gap of 100 bp	
*	29067	29067:	contig of 699 bp in length	
*	29066:	29065:	gap of 100 bp	
*	29066:	29067:	contig of 701 bp in length	
*	29067:	29066:	contig of 712 bp in length	
*	29066:	30577:	contig of 712 bp in length	
*	30578	30677:	gap of 100 bp	
*	30678	31331:	contig of 704 bp in length	
*	31382:	31481:	gap of 100 bp	
*	31482:	32192:	contig of 711 bp in length	
*	32192:	32232:	gap of 100 bp	
*	32232:	32233:	contig of 682 bp in length	
*	32233:	32974:	contig of 682 bp in length	
*	32974:	32975:	gap of 100 bp	
*	32975:	33074:	contig of 686 bp in length	
*	33075:	33076:	gap of 100 bp	
*	33076:	33751:	contig of 714 bp in length	
*	33751:	33860:	gap of 100 bp	
*	33860:	34574:	contig of 682 bp in length	
*	34574:	34575:	gap of 100 bp	
*	34575:	35375:	contig of 704 bp in length	
*	35375:	35478:	gap of 100 bp	
*	35478:	355379:	contig of 686 bp in length	
*	355379:	36194:	contig of 716 bp in length	
*	36194:	36294:	gap of 100 bp	
*	36294:	36973:	contig of 679 bp in length	
*	36973:	37073:	gap of 100 bp	
*	37073:	37774:	contig of 701 bp in length	
*	37774:	37775:	gap of 100 bp	
*	37775:	38586:	contig of 712 bp in length	
*	38586:	38875:	gap of 100 bp	
*	38875:	38587:	contig of 716 bp in length	
*	38587:	39388:	contig of 702 bp in length	
*	39388:	39488:	gap of 100 bp	
*	39488:	40195:	contig of 707 bp in length	
*	40195:	40295:	gap of 100 bp	
*	40295:	41013:	contig of 718 bp in length	
*	41013:	41014:	gap of 100 bp	
*	41014:	41113:	contig of 715 bp in length	
*	41113:	41114:	gap of 100 bp	
*	41114:	41828:	contig of 715 bp in length	
*	41828:	41928:	gap of 100 bp	
*	41928:	42643:	contig of 715 bp in length	
*	42643:	42743:	gap of 100 bp	
*	42743:	43458:	contig of 715 bp in length	
*	43458:	43459:	gap of 100 bp	
*	43459:	43558:	contig of 684 bp in length	
*	43558:	44242:	contig of 715 bp in length	
*	44242:	44342:	gap of 100 bp	
*	44342:	44343:	contig of 701 bp in length	
*	44343:	45043:	contig of 716 bp in length	
*	45043:	45143:	gap of 100 bp	
*	45143:	45244:	contig of 715 bp in length	
*	45244:	45345:	gap of 100 bp	
*	45345:	45856:	contig of 713 bp in length	
*	45856:	48258:	contig of 687 bp in length	
*	48258:	50773:	gap of 100 bp	
*	50773:	51472:	contig of 699 bp in length	
*	51472:	49070:	contig of 712 bp in length	
*	49070:	49170:	gap of 100 bp	
*	49170:	49886:	contig of 716 bp in length	
*	49886:	49986:	gap of 100 bp	
*	49986:	50673:	contig of 687 bp in length	
*	50673:	50674:	gap of 100 bp	
*	50674:	51473:	contig of 698 bp in length	
*	51473:	51572:	gap of 100 bp	
*	51572:	51573:	contig of 716 bp in length	
*	51573:	52288:	contig of 716 bp in length	
*	52288:	52388:	gap of 100 bp	
*	52388:	53075:	contig of 687 bp in length	
*	53075:	53076:	gap of 100 bp	
*	53076:	53175:	contig of 687 bp in length	
*	53175:	53176:	gap of 100 bp	
*	53176:	53879:	contig of 704 bp in length	

RESULT 9  
 AC139040 AC139040 221524 bp DNA linear ROD 16-JUL-2003  
 LOCUS Mus musculus chromosome 19, clone RP23-292H20, complete sequence.  
 DEFINITION AC139040\_7 GI:32813610  
 VERSION HTG.  
 ACCESSION AC139040\_7  
 KEYWORDS Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCES 1 (bases 1 to 221524)  
 AUTHORS Birren,B., Abouelleil,A., Lander,E., Allen,N.,  
 ANDERSON,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 BOGUSLAVSKIY,L., Boukhgalter,B., Camarata,J., Chang,J., Choopel,Y.,  
 COLLYMORE,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
 DIAZ,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
 FERREIRA,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 GRAHAM,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
 HALL,J., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,  
 KAMAT,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 LINDBLAD-TOH,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,  
 MACDONALD,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
 MELDRIM,J., Meneus,L., Mihowa,T., Mlenga,V., Murphy,T., Naylor,J.,  
 O'NEILL,D., Oliver,J., Peterson,K., Phankhang,P., Pierre,N.,  
 RACHUPKA,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 ROMAN,J., Schauer,S., Schupback,R., Seaman,S., Severv,P., Smith,C.,  
 SPENCER,B., Strange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 TALAMAS,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
 VASSILIEV,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 WYMAN,D., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Jul 16, 2003 this sequence version replaced gi:31249817.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green,P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
 Center: Whitehead Institute / MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Center project name: L29083  
 Center clone name: 292\_H\_20

----- Project Information

FEATURES source  
 1. 221524  
 /organism="Mus musculus"  
 /mol type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="19"  
 /map="19"  
 /clone="RP23-292H20"  
 /clone\_id="RPCI-23 Female Mouse BAC"

misc\_feature  
 1. 7046  
 /note="wgs end extension  
 clone\_end \$P6"  
 unsure  
 134. 1239  
 /note=<30 qual SNGI region  
 repeat\_region  
 complement(1312..1743)  
 /rpt\_family="MUT1H1"  
 complement(1981..2123)  
 /rpt\_family="MUR117"  
 repeat\_region  
 2143..2511  
 /rpt\_family="ORR1D"  
 repeat\_region  
 2897..2955  
 /rpt\_family="CA)n"  
 repeat\_region  
 3063..3114  
 /rpt\_family="(TG)n"  
 repeat\_region  
 3141..3217  
 /rpt\_family="BPA"  
 complement(3235..3402)  
 /rpt\_family="BPA"  
 complement(3447..3584)  
 /rpt\_family="BL\_MM"  
 complement(3456..3690)

VASSILIEV,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (31-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 4 (bases 1 to 221524)  
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choopel,Y.,  
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
 Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
 Hall,J., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,  
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
 Meldrim,J., Meneus,L., Mihowa,T., Mlenga,V., Murphy,T., Naylor,J.,  
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,J., Peterson,K., Phankhang,P., Pierre,N.,  
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severv,P., Smith,C.,  
 Spencer,B., Strange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 16, 2003 this sequence version replaced gi:31249817.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green,P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
 Center: Whitehead Institute / MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Center project name: L29083  
 Center clone name: 292\_H\_20

----- Location/Qualifiers

REFERENCE 1. 221524  
 AUTHORS Birren,B., Abouelleil,A., Lander,E., Allen,N.,  
 ANDERSON,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 BOGUSLAVSKIY,L., Boukhgalter,B., Camarata,J., Chang,J., Choopel,Y.,  
 COLLYMORE,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
 DIAZ,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
 FERREIRA,P., FITZGERALD,M., GAGE,D., GALAGAN,J., GARDYNA,S.,  
 GRAHAM,L., GRAND-PIERRE,N., HAFEZ,N., HAGOPIAN,D., HAGOS,B.,  
 HALL,J., HORTON,L., HUIME,W., ILIEV,I., JOHNSON,R., JONES,C.,  
 KAMAT,A., KARATAS,A., KELLS,C., LANDERS,T., LEVINE,R.,  
 LINDBLAD-TOH,K., LIU,X., LUJ,A., MABBITT,R., MACLEAN,C.,  
 MACDONALD,P., MAJOR,J., MANNING,J., MATTHEWS,C., MCCARTHY,M.,  
 MELDRIM,J., MENEUS,L., MIHOWA,T., MLENGA,V., MURPHY,T., NAYLOR,J.,  
 NGUYEN,C., NICOL,R., NORBU,C., O'CONNOR,T., O'DONNELL,P.,  
 O'NEILL,D., OLIVER,J., PETERSON,K., PHANKHANG,P., PIERRE,N.,  
 RACHUPKA,A., RAMASAMY,U., RAYMOND,C., RETTA,R., RISE,C., ROGOV,P.,  
 ROMAN,J., SCHAUER,S., SCHUPBACK,R., SEAMAN,S., SEVERV,P., SMITH,C.,  
 SPENCER,B., STRANGE-THOMANN,N., STOJANOVIC,N., STUBBS,M.,  
 TALAMAS,J., TESFAYE,S., THEODORE,J., TOPHAM,K., TRAVERS,M.,  
 VASSILIEV,H., VENKATARAMAN,V.S., VIEL,R., VO,A., WILSON,B., WU,X.,  
 WYMAN,D., YOUNG,G., ZAINOUN,J., ZIMMER,A. and ZODY,M.  
 Direct Submission  
 Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 16, 2003 this sequence version replaced gi:31249817.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green,P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
 Center: Whitehead Institute / MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Center project name: L29083  
 Center clone name: 292\_H\_20

----- Location/Qualifiers

REFERENCE 1. 221524  
 AUTHORS Birren,B., Abouelleil,A., Lander,E., Allen,N.,  
 ANDERSON,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 BOGUSLAVSKIY,L., Boukhgalter,B., Camarata,J., Chang,J., Choopel,Y.,  
 COLLYMORE,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
 DIAZ,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
 FERREIRA,P., FITZGERALD,M., GAGE,D., GALAGAN,J., GARDYNA,S.,  
 GRAHAM,L., GRAND-PIERRE,N., HAFEZ,N., HAGOPIAN,D., HAGOS,B.,  
 HALL,J., HORTON,L., HUIME,W., ILIEV,I., JOHNSON,R., JONES,C.,  
 KAMAT,A., KARATAS,A., KELLS,C., LANDERS,T., LEVINE,R.,  
 LINDBLAD-TOH,K., LIU,X., LUJ,A., MABBITT,R., MACLEAN,C.,  
 MACDONALD,P., MAJOR,J., MANNING,J., MATTHEWS,C., MCCARTHY,M.,  
 MELDRIM,J., MENEUS,L., MIHOWA,T., MLENGA,V., MURPHY,T., NAYLOR,J.,  
 NGUYEN,C., NICOL,R., NORBU,C., O'CONNOR,T., O'DONNELL,P.,  
 O'NEILL,D., OLIVER,J., PETERSON,K., PHANKHANG,P., PIERRE,N.,  
 RACHUPKA,A., RAMASAMY,U., RAYMOND,C., RETTA,R., RISE,C., ROGOV,P.,  
 ROMAN,J., SCHAUER,S., SCHUPBACK,R., SEAMAN,S., SEVERV,P., SMITH,C.,  
 SPENCER,B., STRANGE-THOMANN,N., STOJANOVIC,N., STUBBS,M.,  
 TALAMAS,J., TESFAYE,S., THEODORE,J., TOPHAM,K., TRAVERS,M.,  
 VASSILIEV,H., VENKATARAMAN,V.S., VIEL,R., VO,A., WILSON,B., WU,X.,  
 WYMAN,D., YOUNG,G., ZAINOUN,J., ZIMMER,A. and ZODY,M.

/rpt_family="ID_B1"	repeat_region	Best Local Similarity 100.0%; Pred. No. 7.5e-12; Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
3724..3752	/rpt_family="(A)n"	
repeat_region	/rpt_complement(4335..4351)	
repeat_region	/rpt_family="B3A"	
repeat_region	/rpt_family="(TG)n"	
repeat_region	/rpt_complement(4409..4625)	
repeat_region	/rpt_family="B3A"	
repeat_region	/rpt_complement(4701..4799)	
repeat_region	/rpt_family="MIR"	
repeat_region	/rpt_complement(5000..5112)	
repeat_region	5268..531	
repeat_region	/rpt_family="ORR1D"	
repeat_region	5782..5810	
repeat_region	/rpt_family="B4"	
repeat_region	5897..6040	
repeat_region	/rpt_family="B1_MM"	
repeat_region	6447..6481	
repeat_region	/rpt_family="(TC)n"	
repeat_region	6481..6516	
repeat_region	/rpt_family="(CA)n"	
repeat_region	7047..7052	
repeat_region	/note="clone_boundary	
repeat_region	Clone end_S6	
repeat_region	site_Ecoli"	
repeat_region	7196..7246	
repeat_region	/rpt_family="(GAAA)n"	
repeat_region	compT_complement(7364..7656)	
repeat_region	/rpt_family="ORR1C"	
repeat_region	8336..8484	
repeat_region	/rpt_family="B1_MM"	
repeat_region	8968..9042	
repeat_region	/rpt_family="L2"	
repeat_region	compT_complement(9139..9332)	
repeat_region	/rpt_family="MIR"	
repeat_region	compT_complement(9553..9742)	
repeat_region	/rpt_family="B2_Mm2"	
repeat_region	compT_complement(9847..10055)	
repeat_region	/rpt_family="B3"	
repeat_region	11638..11844	
repeat_region	/rpt_family="B3"	
repeat_region	12218..12494	
repeat_region	/rpt_family="B2_Mm2"	
repeat_region	compT_complement(9847..10055)	
repeat_region	/rpt_family="B3"	
repeat_region	14922..14986	
repeat_region	/rpt_family="MIR"	
repeat_region	15105..15218	
repeat_region	/rpt_family="LIME3"	
repeat_region	15345..15509	
repeat_region	/rpt_family="LIMC4"	
repeat_region	compT_complement(15704..15731)	
repeat_region	/rpt_family="ID_B1"	
repeat_region	15732..15849	
repeat_region	/rpt_family="ORR1D"	
repeat_region	15850..15874	
repeat_region	/rpt_family="(TGCGGG)n"	
repeat_region	15875..16100	
repeat_region	/rpt_family="ORR1D"	
repeat_region	compT_complement(16101..16251)	
repeat_region	/rpt_family="ID_B1"	
repeat_region	17476..17513	
repeat_region	/rpt_family="(CA)n"	
repeat_region	17626..17653	
repeat_region	/rpt_family="(CCCA)n"	
repeat_region	18382..18451	
repeat_region	/rpt_family="ID4"	
repeat_region	complement(18592..18730)	
repeat_region	/rpt_family="RMR10A"	
	Query Match	17.2%; Score 45; DB 10; Length 221524;

TITLE JOURNAL

Unpublished Direct Submission

REFERENCE 2 (bases 1 to 243412) misc\_feature /note="wgs contig"  
 AUTHORS Worley, K.C. /note="wgs contig"  
 TITLE Direct Submission  
 JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 ORIGIN 3 (bases 1 to 243412) Rat Genome Sequencing Consortium.

REFERENCE 1 GATCAGGGTGGAGTTGAGGAGCTGTGAGACCAAGACGCCCG 45  
 AUTHORS Direct Submission  
 JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On Nov 19, 2002 this sequence version replaced gi:23269487.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GZSV  
 Center clone name: CH230-157C16  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 220496 bases at least Q40  
 Consensus quality: 222898 bases at least Q30  
 Consensus quality: 224623 bases at least Q20  
 Estimated insert size: 224491; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.edu/docs/Genbank/draft\\_data.html](http://www.hgsc.bcm.edu/docs/Genbank/draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1. 11919: contig of 11919 bp in length  
 \* 11920 12019: gap of unknown length  
 \* 12020 18301: contig of 6282 bp in length  
 \* 18302 18401: gap of unknown length  
 \* 18402 243412: contig of 225011 bp in length.

FEATURES Location/Qualifiers  
 source  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="Caxon:10116"  
 /clone="CH230-157C16"  
 1..1535  
 misc\_feature  
 /note="wgs contig"  
 misc\_feature  
 /note="wgs contig"

misc\_feature 1..179734  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="Ctaxon:7955"  
 /clone="CH211-66I20"  
 /clone\_lib="CHORI-211"  
 1..89039

/note="assembly\_fragment:02405  
 fragment\_chain:1  
 clone\_end:17  
 vector\_side:left"  
**misc\_feature**  
 89140 .169174  
 /note="assembly\_fragment:01570  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:right"  
**misc\_feature**  
 169215 .179734  
 /note="assembly\_fragment:00769"  
**ORIGIN**

Query Match Score 8.8%; Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0; Length 179734;

LOCUS AF076980 3291 bp DNA heat shock protein ClpB (clpB) gene, complete cds.

DEFINITION Streptomyces albus G

ACCESSION AF076980

VERSION GI:4322268

KEYWORDS

SOURCE Streptomyces albus G

ORGANISM Bacteria: Actinobacteria: Actinomycetales: Actinomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 3291)  
 AUTHORS Grandvalt,C., de Crecy-Lagard,V. and Mazodier,P.  
 TITLE The ClpB Atpase of Streptomyces albus G belongs to the HspR heat shock regulon  
 JOURNAL Mol. Microbiol. 31 (2), 521-532 (1999)

MEDLINE 9915755  
 PUBMED 10027969

REFERENCE 2 (bases 1 to 3291)  
 AUTHORS Grandvalt,C., de Crecy-Lagard,V. and Mazodier,P.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JUL-1998) Biochimie Microbiene, Institut Pasteur, 25, rue du Docteur Roux, Paris 75724, France  
 FEATURES source 1..3291  
 /organism="Streptomyces albus G"  
 /mol\_type="Genomic DNA"  
 /db\_xref="taxon:1962"  
 -35'\_signal 257..262  
 repeat\_region 271..289  
 /note="HspR associated inverted repeat; binds HspR  
 repressor"  
 /rpt\_Family="HAIR"  
 /rpt\_type=inverted  
 -10'\_signal 281..286  
 gene 457..3030  
 CDS 457..3030  
 /gene="clpB"  
 /gene="clpB"  
 /note="ATPase"  
 /codon\_start=1  
 /trans\_table=11  
 /product="heat shock protein ClpB"  
 /protein\_id="AAD15985.1"  
 /db\_xref="GI:4322268"  
 /translation="MWAELITRSRDLNALAAITRAVSAGNPDLTPAHILLALLEGODNE  
 NLVDEFLISTEHLIGLIRPTARPRCSGRAPKPEKLEAFONTTRGERRVTPRECOYK  
 ALEXRGTDTEIAERKLDPVGDDETRNVQULSRRTKNNPLJIGPCVGKAVVE  
 GLAQRTVKGDYDPESLDKRLVSLDGMVAGAKTRGEFEERLKTVSEIKSDGQ1V  
 G"

**ORIGIN**

FIDELHTVAGAAADSAMDAAGNMLKPMMLARGELMVGGATTLDYERERIEKDPALERRF  
 QVIVVAEPSEVEDSTAILRGKIGRVEAHVQIAUDSRYTTSRFLPKDIA  
 DIVDEAASSURMEIDSSSPIEDELSRSTVRLQIMBELAJDRETDSASRORIEKLRDIA  
 DRERSCAAHRPWEKEKQSINRVEKLERLDLRLQQAERQAQHGDFDTASKLILGEIPT  
 LERDILWRPAEEBAKDTMVKEAGRIPDADWVSWTGPAGRLEGETOKLRLMEA  
 LGRLLIGQSEAVQAVSADVTRAGIADPRTGAFELFLGPTGVCKTELAKAADFED  
 DERAMIRIDNSBEGEHHSVYLICAPGVGEGQQLTEARRPSYMTLDEVEKA  
 HPGVDELLIQYQDQGRDTQGQFNTLVLTSNLSQYLVGSAPPEBKRRQEA  
 VVRSSEFQPFELNRDGLDULVFEALSDEDEARIAGLQAGLARRIAADRSSLDVTRPAA  
 WLAKEGFDPAYGARPLRLIQTAGDRILAKEVDRGDTVRDVEDGLLVRAE

Query Match Score 8.4%; Best Local Similarity 100.0%; Pred. No. 4.2%; Mismatches 0; Indels 0; Gaps 0; Length 3291;

LOCUS AR063085 1425 bp DNA linear PAT 29-SEP-1999

DEFINITION Sequence 6 from patent US 5844079.

ACCESSION AR063085

VERSION AR063085.1 GI:59990776

KEYWORDS Unknown

SOURCE Unknown

ORGANISM Unknown

REFERENCE 1 (bases 1 to 1425)  
 AUTHORS Ingham,P.W., McMahon,A.P. and Tabin,C.J.  
 TITLE Vertebrate embryonic pattern-inducing proteins, and uses related thereto  
 JOURNAL Patent: US 5844079-A 6 01-DEC-1998;  
 FEATURES Location/Qualifiers 1..1425  
 source /organism="unknown"  
 /mol\_type="unassigned DNA"

**ORIGIN**

Query Match Score 8.0%; Best Local Similarity 100.0%; Pred. No. 14; Mismatches 0; Indels 0; Gaps 0; Length 1425;

LOCUS AR122612 1425 bp DNA linear PAT 16-MAY-2001

DEFINITION Sequence 6 from patent US 6165747.

ACCESSION AR122612

VERSION AR122612.1 GI:14106929

KEYWORDS Unknown

SOURCE Unknown

ORGANISM Unknown

REFERENCE 1 (bases 1 to 1425)  
 AUTHORS Ingham,P.W., McMahon,A.P., Tabin,C.J., Bumcrot,D.A. and Marti-Gorostiza,E  
 TITLE Nucleic acids encoding hedgehog proteins  
 JOURNAL Patent: US 6165747-A 6 26-DEC-2000;  
 FEATURES Location/Qualifiers 1..1425  
 source /organism="unknown"  
 /mol\_type="unassigned DNA"

**ORIGIN**

Query Match      Score 21; DB 6; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 21; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

Qy    139 CTCGGCCACCGGGAGGCCGA 159  
 Db    837 CTCGGCCACCGGGAGGCCGA 857

---

RESULT 15

AR164240 AR164240 Sequence 6 from patent US 6271363. DNA linear PAT 17-OCT-2001

DEFINITION Sequence 6 from patent US 6271363.

ACCESSION AR164240

VERSION AR164240.1 GI:16235295

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM Unknown.

UNCLASSIFIED

REFERENCE 1. (bases 1 to 1425)

AUTHORS Ingham, P.W., McMahon, A.P. and Tabin, C.J.

TITLE Nucleic acids encoding hedgehog proteins

JOURNAL Patent: US 6271363-A 6 07-AUG-2001;

FEATURES Location/Qualifiers

source 1..1425

/organism="unknown"

/mol\_type="unassigned DNA"

ORIGIN

Query Match      Score 21; DB 6; Length 1425;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 21; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;

Qy    139 CTCGGCCACCGGGAGGCCGA 159  
 Db    837 CTCGGCCACCGGGAGGCCGA 857

Search completed: July 18, 2004, 15:18:48  
 Job time : 1615 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: July 18, 2004, 11:40:25 ; Search time 209.026 Seconds  
(without alignments)  
4583.781 Million cell updates/sec

Title: US-10-051-769-5  
Perfect score: 22  
Sequence: 1 aaggtagttcgaggatgc 22

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 3470272 seqs., 21671516995 residues

Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_dat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rod:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	22	100.0	2623	9 BC012186	BC012186 Homo sapi
2	22	100.0	2639	6 AX741159	AX741159 Sequence
3	22	100.0	2639	9 AK091746	AK091746 Homo sapi
4	22	100.0	3856	9 BC028375	BC028375 Homo sapi
C	5	22	100.0	49052	9 AL359836 Human DNA
C	6	22	100.0	68195	2 AC108407 Mus muscu
C	7	22	100.0	120173	AC005887 c1t0_173
				10 AC139040	Mus muscu
				2 AC127769	Rattus no
				10 AC105446	Homo sapi
				11 AL539126	Streptomy
				12 AF110372	Leishmani
				13 MG3109	Leishmani
				14 AX98410	
				15 AF051720	Drosophil
				16 AY059441	Drosophil
				17 U98086	Hydrogenop
				18 AF489516	Methylloba
				19 AB040942	Homo sapi
				20 AX198410	
				21 AF169049	Gallus ga
				22 AC014121	Drosophil
				23 AC026266	Homo sapi
				24 AC012693	Drosophil
				25 AC018520	Continuation (3 of
				26 AC018520	Homo sapi
				27 AC0097347	Homo sapi
				28 AC009216	Drosophil
				29 AC012162	Drosophil
				30 AC004394	Homo sapi
				31 AC003470	Drosophil
				32 AC003470	Drosophil
				33 AC01127	Drosophil
				34 AC013507	Homo sapi
				35 AC003470	Drosophil
				36 AC003470	Drosophil
				37 AC003470	Drosophil
				38 AC003470	Drosophil
				39 AC014916	Drosophil
				40 AC1158043	Human DNA
				41 AC010040	Drosophil
				42 AC010956	Homo sapi
				43 AC125957	Rattus no
				44 AC129582	Mus muscu
				45 AL596182	Mouse DNA

## ALIGNMENTS

RESULT 1	BC012186	2623 bp	mRNA	PRI 06-AUG-2001
LOCUS	Homo sapiens, clone BC012186	IMAGE:4564853,	mRNA	
DEFINITION				
VERSION	BC012186.1	GI:1502555		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	1 (bases 1 to 2623)			
AUTHORS	Strausberg,R.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-AUG-2001) National Institutes of Health, Mammalian			

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DTP  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILN)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcgsc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Choi, Chris Fjell, Erin Garland, Ran Guin, Letricia Hsiao, Martin Krzyniowski, Reta Kutsche, Oliver Lee, See Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILN at: <http://image.liln.gov>  
 Series: IRL Plate: 29 Row: b Column: 17  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomeScan gene prediction.

#### FEATURES

source  
 /organism="Homo sapiens"  
 /db\_xref="IMAGE:9606"  
 /clone\_type="mRNA"  
 /note="IMAGE:4564833"  
 /tissue\_type="Kidney, renal cell adenocarcinoma"  
 /clone\_id="NIH MGC\_14"  
 /lab\_host="DH10B-R"  
 /note="Vector: POTB?"

#### ORIGIN

Query Match Score 22; DB 9; Length 2623;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AACGTGGAGTTGAGGAGCTGC 22  
 Db 547 AAGGTGGAGTTGAGGAGCTGC 568

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Maresu,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Kawiyama,A., Nagai,K., Masuhira,Y., Sugano,S., Nagahara,K., Matsuo,Y., Nagai,K. and Isogai,T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 2639)

#### REFERENCE

Isogai,T. and Yamamoto,J.  
 Direct Submission  
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 JOURNAL Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3975, Fax: 81-438-52-3986  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'-&-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB;  
 RAB; annotation: HRI and RAB.

#### FEATURES

##### SOURCE

1..2639  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_id="HHDPC200816"  
 /cell\_type="dermal papilla cells (HDDPC)"  
 /clone\_idb="HHDPC22"  
 /note="cloning vector: pME18SFL3-primary culture, dermal papilla cells"

#### ORIGIN

Query Match Score 22; DB 9; Length 2639;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAGGTGGAGTTGAGGAGCTGC 22  
 Db 583 AAGGTGGAGTTGAGGAGCTGC 604

#### REFERENCE

Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Saito,K., Nishikawa,T., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagai,K., and Masuhira,Y.  
 Full-length cDNA sequences  
 Patent: EP-1308459-A 684 07-MAY-2003;  
 Helix Research Institute (JP); Research Association for Biotechnology (JP)  
 Location/Qualifiers  
 source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

#### FEATURES

##### SOURCE

1..2639  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

#### ORIGIN

BC028375 mRNA linear PRI 07-OCT-2003  
 Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone MGC:27107 IMAGE:4837939), complete cds.  
 BC028375

ORIGIN  
 Query Match Score 22; DB 6; Length 2639;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AACGTGGAGTTGAGGAGCTGC 22  
 Db 583 AAGGTGGAGTTGAGGAGCTGC 604

#### REFERENCE

AK091746 mRNA linear PRI 15-JUL-2002  
 Homo sapiens cDNA FLJ34427 Fis, clone HHDPC200816.  
 DEFINITION Homo sapiens cDNA FLJ34427 Fis, clone HHDPC200816.  
 ACCESSION AK091746  
 VERSION AK091746.1  
 KEYWORDS oligoapping, fis (full insert sequence).  
 SOURCE Homo sapiens (human).  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

#### REFERENCE

Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Maresu,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Kawiyama,A., Nagai,K., Sugano,S., Nagahara,K., Masuhira,Y., Nagai,K. and Isogai,T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 1..2639  
 Isogai,T. and Yamamoto,J.  
 Direct Submission  
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 JOURNAL Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3975, Fax: 81-438-52-3986  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'-&-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB;  
 RAB; annotation: HRI and RAB.

#### REFERENCE

Location/Qualifiers  
 source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_id="HHDPC200816"  
 /cell\_type="dermal papilla cells (HDDPC)"  
 /clone\_idb="HHDPC22"  
 /note="cloning vector: pME18SFL3-primary culture, dermal papilla cells"

#### REFERENCE

1..2639  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

#### REFERENCE

BC028375 mRNA linear PRI 07-OCT-2003  
 Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone MGC:27107 IMAGE:4837939), complete cds.  
 BC028375



an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL; EMBL; SWISSPROT; Tr.; TREMBL; Tr.; WORMPPE; Information on WORMPPE database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormppe](http://www.sanger.ac.uk/Projects/C_elegans/wormppe). This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrl0>. RP11-389B6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>.

RELOCATOR: pbACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-389B6. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CPA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

FEATURES	source	Query	Match	Score	DB	Length	Qualifiers
	source	1.	4.9052	100.0%	DB	9;	Length 49052;
	/organism="Homo sapiens"				No.	23;	
	/mol type="genomic DNA"						
	/db_xref="taxon:9606"						
	/chromosome="10"						
	/clone="RP11-389E6"						
	/clone_lname="RP01-11.2"						
ORIGIN							
		Query	1	AAGGTGGAGGTGCGGGAGCTGC	22		
			29763	AAGGTGGAGGTGCGGGAGCTGC	29742		
RESULT	AC108407			68196 bp	DNA	linear	HTG 27-JAN-2000
DEFINITION	Mus musculus clone RP24-422P10,						LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC108407						
VERSION	AC108407.1						
KEYWORDS	HTG; HTGS PHASE0;						
SOURCE	Mus musculus (house mouse)						
ORGANISM	Mus musculus						
MATERIALS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.						
REFERENCE	1 (bases 1 to 68196)						
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.						
TITLE	Mus musculus, clone RP24-422P10						
JOURNAL	Unpublished						
REFERENCE	2 (bases 1 to 68196)						
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguski, V.L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeAngelis, S., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Haigis, B., Horion, L., Huike, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelly, C., LaRoque, K., Lamazares R., Landers, T., Lebockzy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Margolis, J., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Medrano, J., Menes, L., Mihova, T., Milanga, V., Murphy, T., Naylor, J.J., Nguyen, C., Nicol, R., Nobu, C., Norman, C.H., O'Connor, T., P. O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,						

---

* * 16144	16859; contiguous of 716 bp in length	*	45857	45955; gap of 100 bp
16860	gap of 100 bp	*	45957	contig of 638 bp in length
* * 16960	17669; contiguous of 710 bp in length	*	46654; gap of 100 bp	
* * 17670	17763; gap of 100 bp	*	46755; gap of 100 bp	
* * 17770	18460; contiguous of 691 bp in length	*	47449; gap of 635 bp in length	
* * 18461	18560; gap of 100 bp	*	47559; gap of 100 bp	
* * 18561	19269; contiguous of 709 bp in length	*	48258; contiguous of 709 bp in length	
* * 19270	19369; gap of 100 bp	*	48259; gap of 100 bp	
* * 19370	20070; contiguous of 701 bp in length	*	48354; gap of 100 bp	
* * 20071	20170; gap of 100 bp	*	49070; contiguous of 712 bp in length	
* * 20171	20874; contiguous of 704 bp in length	*	49170; gap of 100 bp	
* * 20875	20974; gap of 100 bp	*	49884; contiguous of 716 bp in length	
* * 20975	21684; contiguous of 710 bp in length	*	49985; gap of 100 bp	
* * 21685	21754; gap of 100 bp	*	49987; contiguous of 687 bp in length	
* * 21785	22491; contiguous of 707 bp in length	*	50673; gap of 100 bp	
* * 22492	22591; gap of 100 bp	*	51472; contiguous of 693 bp in length	
* * 22592	23292; contiguous of 708 bp in length	*	51473; gap of 100 bp	
* * 23300	23399; gap of 100 bp	*	52288; contiguous of 716 bp in length	
* * 23400	24086; contiguous of 687 bp in length	*	52389; gap of 100 bp	
* * 24087	24186; gap of 100 bp	*	53075; contiguous of 687 bp in length	
* * 24187	24901; contiguous of 715 bp in length	*	53076; gap of 100 bp	
* * 24902	25001; gap of 100 bp	*	53875; contiguous of 704 bp in length	
* * 25002	25714; contiguous of 713 bp in length	*	53876; gap of 100 bp	
* * 25715	25814; gap of 100 bp	*	53979; gap of 100 bp	
* * 25815	26528; contiguous of 714 bp in length	*	54675; contiguous of 636 bp in length	
* * 26529	26628; gap of 100 bp	*	54676; gap of 100 bp	
* * 26629	27358; contiguous of 730 bp in length	*	54776; contiguous of 702 bp in length	
* * 27359	27458; gap of 100 bp	*	55478; gap of 100 bp	
* * 27459	28159; contiguous of 701 bp in length	*		
* * 28160	28259; gap of 100 bp	*		
* * 28260	28966; contiguous of 707 bp in length	*		
* * 28967	29066; gap of 100 bp	*		
* * 29067	29765; contiguous of 699 bp in length	*		
* * 29766	29865; gap of 100 bp	*		
* * 29866	30577; contiguous of 712 bp in length	*		
* * 30578	30677; gap of 100 bp	*		
* * 30678	31381; contiguous of 704 bp in length	*		
* * 31382	31481; gap of 100 bp	*		
* * 31482	32192; contiguous of 711 bp in length	*		
* * 32193	32292; gap of 100 bp	*		
* * 32293	32974; contiguous of 682 bp in length	*		
* * 32975	33074; gap of 100 bp	*		
* * 33075	33760; contiguous of 686 bp in length	*		
* * 33761	33860; gap of 100 bp	*		
* * 33861	34574; contiguous of 714 bp in length	*		
* * 34575	34674; gap of 100 bp	*		
* * 35378	35378; contiguous of 704 bp in length	*		
* * 35379	35478; gap of 100 bp	*		
* * 35479	36194; contiguous of 716 bp in length	*		
* * 36195	36294; gap of 100 bp	*		
* * 36295	36973; contiguous of 679 bp in length	*		
* * 36974	37073; gap of 100 bp	*		
* * 37074	37774; contiguous of 701 bp in length	*		
* * 37775	37874; gap of 100 bp	*		
* * 37875	38586; contiguous of 712 bp in length	*		
* * 38587	38686; gap of 100 bp	*		
* * 38687	39388; contiguous of 702 bp in length	*		
* * 39389	39488; gap of 100 bp	*		
* * 39489	40195; contiguous of 707 bp in length	*		
* * 40196	40295; gap of 100 bp	*		
* * 40296	41013; contiguous of 718 bp in length	*		
* * 41014	41113; gap of 100 bp	*		
* * 41114	41828; contiguous of 715 bp in length	*		
* * 41829	41928; gap of 100 bp	*		
* * 41929	42643; contiguous of 715 bp in length	*		
* * 42644	42743; gap of 100 bp	*		
* * 42744	43458; contiguous of 715 bp in length	*		
* * 43459	43558; gap of 100 bp	*		
* * 43559	44242; contiguous of 684 bp in length	*		
* * 44243	44342; gap of 100 bp	*		
* * 44343	45043; contiguous of 701 bp in length	*		
* * 45044	45143; gap of 100 bp	*		
* * 45144	45856; contiguous of 713 bp in length	*		

Query Match  
Best Local Similarity 100.0%; Score 22; DB 2; Length 68196;  
Matches 22; Conservative 0; Mismatches 0; Gaps 0;

Qy	1 AAGGTGGAGTTGAGGGCTGC 22	
Db	17501 AAGGTGGAGTTGAGGGCTGC 17480	
RESULT 7		
AC005887		
LOCUS	Citb_173_i_12, complete sequence.	
DEFINITION		
ACCESSION	AC005887	
VERSION	AC005887.3	
KEYWORDS	GT:6249675 HTG.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Fukuyama, Metrazo, Chordata, Craniata, Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		
AUTHORS	1 (bases 1 to 120578)	
TITLE	Sequencing of Human Chromosome 10	
JOURNAL	Unpublished	
REFERENCE		
AUTHORS	2 (bases 1 to 120578)	
TITLE	Smith, D.R.	
JOURNAL	Direct Submission	
REFERENCE		
AUTHORS	3 (bases 1 to 120578)	
TITLE	Smith, D.R.	
JOURNAL	Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA	
REFERENCE		
AUTHORS	4 (bases 1 to 120578)	
TITLE	Smith, D.R.	
JOURNAL	Direct Submission	
REFERENCE		
AUTHORS	5 (bases 1 to 120578)	
TITLE	Smith, D.R.	
JOURNAL	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA	
REFERENCE		
AUTHORS	REMARK	
TITLE	Sequence Clipped	
JOURNAL	On Nov 5, 1999 this sequence version replaced gi:4314331.	
FEATURES	Location/Qualifiers	



```

repeat_region complement(3447..3584)
/rpt_feature Family="B1_MM"
complement(ID_B1" 3456..3690)
repeat_region /rpt_feature Family="ID_B1"
repeat_region 3724..3752 /rpt_feature "(A)n"
repeat_region complement(4335..4351)
repeat_region 4352..4408 /rpt_feature Family="B3A"
repeat_region complement("TG)n"
repeat_region complement(4409..4625)
repeat_region complement(4711..4799)
repeat_region complement(5000..5112)
repeat_region /rpt_feature Family="B1F"
repeat_region 5268..5431 /rpt_feature "ORR1D"
repeat_region 5782..5870 /rpt_feature "B4"
repeat_region 5897..6040 /rpt_feature Family="B1_MM"
repeat_region 6447..6481 /rpt_feature "(TC)n"
repeat_region 6481..6516 /rpt_feature "(CA)n"
repeat_region 7047..7052 /rpt_feature Family="B1_MM"
note="clone_boundary
clone_end:SP6
site:EcORI"
repeat_region 7196..7246 /rpt_feature Family="(GAAA)n"
repeat_region complement(7364..7656)
repeat_region 8336..8484 /rpt_feature "ORR1C"
repeat_region 8968..9042 /rpt_feature "B1_MM"
repeat_region 9042..9139 /rpt_feature "ML2"
repeat_region complement(9139..9332)
repeat_region /rpt_feature Family="MIR"
repeat_region complement(9553..9742)
repeat_region /rpt_feature Family="B2_Mm2"
repeat_region complement(9847..10055)
repeat_region /rpt_feature Family="B3"
repeat_region 11638..11844 /rpt_feature Family="B3"
repeat_region 12218..12494 /rpt_feature "RMR14"
repeat_region complement(13731..13993)
repeat_region /rpt_feature Family="B4"
repeat_region 14922..14986 /rpt_feature Family="MIR"
repeat_region 15105..15211 /rpt_feature Family="LIME3"
repeat_region 15345..15509 /rpt_feature Family="LIME4"
repeat_region complement(15704..15731)
repeat_region /rpt_feature Family="ID_B1"
repeat_region /rpt_feature Family="ORR1D"
repeat_region 15850..15874 /rpt_feature Family="LIMC4"
repeat_region complement(15704..15731)
repeat_region /rpt_feature Family="ID_B1"
repeat_region /rpt_feature Family="ORR1D"
repeat_region 17479..17513 /rpt_feature Family="B1"
repeat_region /rpt_feature Family="B4"
repeat_region complement("TGGGG)n"
repeat_region 15875..16100 /rpt_feature Family="MIR"
repeat_region complement(16101..16251)
repeat_region /rpt_feature Family="ID_B1"
repeat_region 18382..18451 /rpt_feature Family="ID4"
repeat_region complement(18592..18730)

```

---

/ rpt_feature Family="RMR10A"	
Query Match	100.0%
Best Local Matches	100.0%
Similarity	100.0%
Conservative	0
Mismatches	0
Indels	0
Gaps	0
	;
Qy	1 AAGGTGAGTCGAGGAGCTGC 22
Db	42837 AAGGTGAGTCGAGGAGCTGC 42858
RESULT	9
AC127769	AC127769
LOCUS	Rattus norvegicus clone CH23-0-157C16, WORKING DRAFT SEQUENCE, 3
DEFINITION	Rattus norvegicus clone CH23-0-157C16, WORKING DRAFT SEQUENCE, 3
ACCESSION	AC127769
VERSION	GI:25079525
KEYWORDS	HTGS PHASEI; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 243412)
AUTHORS	Muzny D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalobechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blith,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,R., Chiu,J., Cleveland,C., Cockrell,I.R., Cox,C., Coyne,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Demayo,C., Ding,Y., Dinh,H., Divya,X., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escott,C., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garcia,M., Gebrgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guerara,W., Gunaratne,P., Haaland,W., Hamilson,C., Hamilton,C., Harvey,Y., Havlik,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollands,B., Howell,S., Hulky,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Khan,S., King,L., Kovar,C., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Louisseged,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindatre,M., Malmoud,M., Mallory,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meinen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Nankervis,C., Neal,D., Newton,A., Pal,S., Parks,K., Nwakelend,O., Okuyon,G., Olatunbosogon,A., Perez,L., Pfannkoch,C., Pasternak,S., Paul,H., Perez,A., Pineda,E., Pu,L.-L., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scheer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sned,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steinle,M., Strong,R., Sutton,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wileczky,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Smith,H.O.,

TITLE Weinstock, G. and Gibbs, R.A.  
 JOURNAL Direct Submission  
 REFERENCE Unpublished  
 2. (bases 1 to 243412)  
 AUTHORS Worley, K.C.  
 JOURNAL Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3. (bases 1 to 243412)  
 REFERENCE Rat Genome Sequencing Consortium.  
 AUTHORS JOURNAL Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Nov 19, 2002 this sequence version replaced qj:23269487.  
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a contig-scaffold). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: GZSV  
 Center clone name: CH230-157C16  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus Quality: 220496 bases at least Q40  
 Consensus Quality: 222898 bases at least Q30  
 Consensus Quality: 224623 bases at least Q20  
 Estimated insert size: 224491; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length.  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank/draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 \* 11919: contig of 11919 bp in length  
 \* 11920 12019: gap of unknown length  
 \* 12020 18301: contig of 6282 bp in length  
 \* 18302 18401: gap of unknown length  
 \* 18402 243412: contig of 225011 bp in length.  
 Location/Qualifiers  
 1. .243412

FEATURES source  
 misc\_feature  
 misc\_feature  
 misc\_feature  
 misc\_feature

misc\_feature 25075. .27269  
 /note="wgs contig"  
 misc\_feature 23839. .242288  
 /note="wgs contig"  
 misc\_feature 243339. .243412  
 /note="wgs contig"

ORIGIN Query Match 100.0%; Score 22; DB 2; Length 243412;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGTGGAGTCGAGAGCTGC 22  
 Db 90934 AAGTGGAGTCGAGAGCTGC 90855

RESULT 10  
 AC105446 139517 bp DNA linear  
 LOCUS Homo sapiens BAC clone RP11-395B7 from 7, complete sequence.  
 DEFINITION Homo sapiens BAC clone RP11-395B7 from 7, complete sequence.  
 AC105446 AC105446.4 GI:19071695  
 VERSION HTG  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 139517)  
 AUTHORS Sulston, J.E. and Wilson, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99053792  
 PUBLISHED 984/10/4  
 REFERENCE 2 (bases 1 to 139517)  
 AUTHORS Doeberl, A., Kalicki, J. and Meyer, R.  
 TITLE The sequence of Homo sapiens BAC clone RP11-395B7  
 JOURNAL Unpublished (2001)  
 MEDLINE 3 (bases 1 to 139517)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 MEDLINE 4 (bases 1 to 139517)  
 REFERENCE 5 (bases 1 to 139517)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-MAR-2002) Genome Sequencing Center, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 MEDLINE 6 (bases 1 to 139517)  
 REFERENCE 5 (bases 1 to 139517)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-APR-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 MEDLINE 7 (bases 1 to 139517)  
 REFERENCE 7 (bases 1 to 139517)  
 AUTHORS Wilson, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-OCT-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 MEDLINE 8 (bases 1 to 139517)  
 COMMENT On Mar 2, 2002 this sequence version replaced gi:18370045.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: sapiens@watson.wustl.edu  
 ----- Summary Statistics  
 Center project name: NH0395B07

9748 . 10043  
repeat\_region /rpt\_family="Alu"  
repeat\_region 10702 . 10781  
repeat\_region /rpt\_family="M4LR"  
repeat\_region 10782 . 10918  
repeat\_region /rpt\_family="Alu"  
repeat\_region 10919 . 10983  
repeat\_region /rpt\_family="M4LR"  
repeat\_region 11012 . 11106  
repeat\_region /rpt\_family="Alu"  
repeat\_region 11110 . 11193  
repeat\_region /rpt\_family="Alu"  
repeat\_region 11145 . 11175  
repeat\_region /rpt\_family="L1"  
repeat\_region 11176 . 112103  
repeat\_region /rpt\_family="Alu"  
repeat\_region 12104 . 12150  
repeat\_region /rpt\_family="L1"  
repeat\_region 13244 . 13350  
repeat\_region /rpt\_family="Alu"  
repeat\_region 13562 . 13853  
repeat\_region /rpt\_family="Alu"  
repeat\_region 13855 . 14010  
repeat\_region /rpt\_family="MIR"  
repeat\_region 14011 . 14318  
repeat\_region /rpt\_family="Alu"  
repeat\_region 14319 . 14417  
repeat\_region /rpt\_family="MIR"  
repeat\_region 14434 . 14608  
repeat\_region /rpt\_family="MIR"  
repeat\_region 14609 . 14901  
repeat\_region /rpt\_family="Alu"  
repeat\_region 14959 . 15105  
repeat\_region /rpt\_family="Alu"  
repeat\_region 15132 . 15377  
repeat\_region /rpt\_family="M4LR"  
repeat\_region 15654 . 15758  
repeat\_region /rpt\_family="MIR"  
repeat\_region 16056 . 16135  
repeat\_region /rpt\_family="Alu"  
repeat\_region 16668 . 16730  
repeat\_region /rpt\_family="(TAA)n"  
repeat\_region 16732 . 16737  
repeat\_region /rpt\_family="(TA)n"  
repeat\_region 16738 . 17039  
repeat\_region /rpt\_family="Alu"  
repeat\_region 17040 . 17059  
repeat\_region /rpt\_family="Alu"  
repeat\_region 17112 . 17261  
repeat\_region /rpt\_family="Alu"  
repeat\_region 17269 . 17579  
repeat\_region /rpt\_family="Alu"  
repeat\_region 17600 . 17879  
repeat\_region /rpt\_family="L1"  
repeat\_region 17880 . 18169  
repeat\_region /rpt\_family="Alu"  
repeat\_region 18170 . 18177  
repeat\_region /rpt\_family="L1"  
repeat\_region 19105 . 19404  
repeat\_region /rpt\_family="Alu"  
repeat\_region 19663 . 19966  
repeat\_region /rpt\_family="Alu"  
repeat\_region 19967 . 20272  
repeat\_region /rpt\_family="L1"  
repeat\_region 20307 . 20375  
repeat\_region /rpt\_family="MIR"  
repeat\_region 20376 . 20675  
repeat\_region /rpt\_family="Alu"  
repeat\_region 20676 . 20706  
repeat\_region /rpt\_family="MIR"  
repeat\_region 20814 . 20889  
repeat\_region /rpt\_family="L2"  
repeat\_region 21157 . 21234

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NIGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.ncbi.nih.gov/DIR/GTB/CHR7>, send mailto:[egreen@nigri.nih.gov](mailto:egreen@nigri.nih.gov) , or see <http://genome.wustl.edu>

**SOURCE INFORMATION:**  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

**VECTOR:** pBACE3.6.  
**Location/Qualifiers**

source	feature	start	end	strand	label
	1 . 139517				/organism="Homo sapiens"
					/mol_type="genomic DNA"
					/db_xref="taxon:9606"
					/chromosome="7"
					/map="7"
					/clone_id="RPCI-11"
					/clone_lib="RPCI-11"
					290 . 591
					/rpt_family="Alu"
					1071 . 1105
					/rpt_family="(GA)n"
					1159 . 1224
					/rpt_family="Alu"
					1275 . 1312
					/rpt_family="(TAA)n"
					1313 . 1471
					/rpt_family="Alu"
					2694 . 3750
					/note="#CPG Island (%GC=69.7, o/e=0.77, #CPGs=89)"
					3678 . 3700
					/rpt_family="GC-rich"
					3917 . 4005
					/rpt_family="CT-rich"
					4452 . 4587
					/rpt_family="MIR"
					5342 . 5438
					/rpt_family="C-rich"
					7443 . 7827
					/rpt_family="Alu"
					7843 . 7973
					/rpt_family="Alu"
					7974 . 8244
					/rpt_family="Alu"
					8245 . 8418
					/rpt_family="Alu"
					9415 . 9501
					/rpt_family="MIR"

Query Match 88.2%; Score 19.4; DB: 9; Length 139517;  
 Best Local Similarity 95.2%; Pred. No. 3.8e+01; Mismatches 0; Gaps 0;  
 Matches 20; Conservative 0;

Qy 1 AAGGTGGAGTTCGAGGAGCTG 21  
 Db 104373 AAGGTGGAGTTCGAGGAGCTG 104393

RESULT 11  
 SC09126 SC0939126 295150 bp DNA linear BCT 11-FEB-2003  
 LOCUS Streptomyces coelicolor A3 (2) complete genome; segment 23/29  
 DEFINITION AL035126 AL035119 AL0323717 AL031514 AL03446 AL0344292  
 ACCESSION AL035212 AL035478 AL049727 AL132824 AL132997 AL598708 AL591522  
 VERSION 1.0  
 KEYWORDS  
 SOURCE ORGANISM Streptomyces coelicolor A3 (2)  
 Bacteria; Actinobacteria; Streptomyctaceae; Streptomyces; Streptomyces coelicolor A3 (2);  
 Streptomyctaceae; Streptomyces; Streptomyctaceae; Streptomyces.

REFERENCE 1  
 AUTHORS Bentley, S.D., Chater, K.P., Cerdeno-Taraga, A.M., Challis, G.L.,  
 Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,  
 Barter, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,  
 Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howard, S.,  
 Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S.,  
 Rabinowitzsch, E., Rajandream, M.A., Rutherford, K., Rutter, S.,  
 Seeger, K., Saunders, D., Sharp, S., Squares, S.,  
 Taylor, K., Warren, T., Wetzorek, A., Woodward, J., Barrell, B.G.,  
 Parkhill, J., and Hopwood, D.A.

TITLE Complete genome sequence of the model actinomycete Streptomyces coelicolor A3 (2)

JOURNAL Nature 417 (6885), 141-147 (2002)

MEDLINE 21936410  
 PUBMED 212000953  
 REFERENCE 2 (bases 1 to 295150)  
 AUTHORS Bentley, S.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxtion, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
 COMMENT gi:20520744, gi:20520662, gi:20520663, gi:20520667, gi:20520669, On or before Oct 26, 2002 this sequence version replaced  
 gi:20520761, gi:20520616, gi:20520670, gi:20520820, gi:20520846,  
 FEATURES source  
 /organism="Streptomyces coelicolor A3 (2)"  
 /mol\_type="genomic DNA"  
 /strain="A3 (2)"  
 /db\_xref="taxon:100226"  
 RBS 512..515  
 /note="possible RBS upstream of SC7B7.02"  
 gene 524..1933  
 /gene="SC06005"  
 /note="SC7B7.02" synonym: SC7B7.02"  
 CDS 524..1933  
 /gene="SC06005"  
 /note="SC7B7.02", possible lipoprotein, len: 469; has similarity to signal sequence of Mycobacterium leprae (445 aa), fasta scores; opt: 75; z-score: 199.2 E(); 0.00062, 23.3% identity in 335 aa overlap. Contains N-terminal signal sequence and appropriately positioned P5000013 Prokaryotic membrane lipoprotein site"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative lipoprotein"  
 /protein\_id="CAA15782.1"

/db\_xref="GI:2661677"  
 /db\_xref="GOA:050499"  
 /note="SPTRMBL:050499"  
 /transl\_table="MTIAGSLDERTLRLGAIATAAMSFAVACSSPSPBDKBDGPPKGERSANPFGAAANSTVAAIFPGYGTIDYDANOVLGSOYKGLKTYPKVVDIAPIQQLPERFVGMPFPDIILSGCEDQIGFLGILDQLEPEASFYEGKTAADIVYPGVKDPGTFKDKEVALNYTMVWYSLTFLPEENGWTPPKMDEALDLQGEARKKKYLVVHGKBAATYTTLJDSAKEGGDDEVLALLENLKGCMWPAVOCVIMKETMTYKORMFVGCGSGTQQKAQIAINSDQKALYPSQGKQVWYSLTFLPEENGWTPPKMDEALDLQGEARKKKYLVVHGKBAATYTTLJDSAKEGGDDEVLALLENLKGCMWPAVOCVIMKETMTYKORMFVGCGSTALVLSQTKMLEACTAQNLYMFETYGLNTDQLYPWNNSPLAGLDGKGHTSALQKISDKVREDDSVKVKVS"  
 581..613  
 /gene="SC06005"  
 /note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"  
 1941..3002  
 /gene="SC06006"  
 /note="synonym: SC7b7.03"  
 1941..3002  
 /gene="SC06006"  
 /note="SC7b7.03, probable integral membrane transport protein, len: 353 aa; similar to many members of the MALFGG subfamily of the binding-protein-dependent transport system membrane proteins eg. MSMF\_STRM\_Qu0750 multiple sugar-binding transport system permease protein (290 aa), fasta scores: opt: 393; z-score: 636.8 E(); 2.6e-28, 32.5% identity in 280 aa overlap. Contains PS00402\_Binding-protein-dependent transport systems inner membrane comp signature"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable transmembrane transport protein"  
 /protein\_id="CAA15783.1"  
 /db\_xref="GI:2661678"  
 /db\_xref="GOA:050500"  
 /db\_xref="SPTRMBL:050500"  
 /note="PS00402\_Binding-protein-dependent transport systems inner membrane component signature"  
 2982..2982  
 /gene="SC06006"  
 /note="possible RBS upstream of SC7b7.04"  
 2999..3907  
 /gene="SC06007"  
 /note="synonym: SC7b7.04"  
 2999..3907  
 /gene="SC06007"  
 /note="SC7b7.04, probable integral membrane transport protein, len: 301 aa; similar to many members of the MALFGG subfamily of the binding-protein-dependent transport systems membrane proteins eg. TR:Q06111 (EMBL:X66692). Putative transport system permease protein from Clostridium perfringens (275 aa), fasta scores: opt: 590 z-score: 796.5 E(); 0, 33.6% identity in 265 aa overlap. Contains PS00402\_Binding-protein-dependent transport systems inner membrane comp signature"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable transmembrane transport protein"  
 /protein\_id="CAA15784.1"  
 /db\_xref="GI:2661679"  
 /db\_xref="GOA:050501"  
 /db\_xref="SPTRMBL:050501"  
 /translation="MSVIAADRKPASDRKEVSDRKLARAASSDRRAISHALLIW"

SVTIVTPMLWVIMSSPFKSTGPBILSSPFLPDHWRFENYANAWTDANIGKFVLNSVIVV VSALILVMLGAMCAYVLARPEFPERLIIYVMMIAGLTETVPLAIVPLFFOLQNGLL NTRGLLPHIFPLFLSFLSPFLPHDVEALIDAGWRAFFVMLMPARP GMAVATIAENFLGLWWOFLFLPVALNTQDKWVLTQMAASSQYDIDGALFAAVV TVVPVLLVYCVRRIAGSVSQGTFR"	Db	233112	GCTGGAGTTTCGAGGAGCTG	233130
misc_feature		RESULT 12		
stem_loop		LOCUS	Leishmania major clone LmV5-G1 putative glycoprotein gene, partial	
RBS		DEFINITION	cds.	
gene		ACCESSION	AF110372	
		VERSION	AF110372.1	GI:4731175
		KEYWORDS	Leishmania major	
		ORGANISM	Leishmania major	
		Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;		
		Leishmania.		
		REFERENCE	1 (bases 1 to 274)	
		AUTHORS	Ch'ang,L.Y. and Li,T.	
		JOURNAL	Trapping ORF-containing genomic fragments	
		REMARKS	Unpublished	
		2 (bases 1 to 274).		
		REMARKS	Ch'ang,L.Y. and Li,T.	
		REMARKS	Direct Submission	
		REMARKS	Submitted (01-DEC-1998) Academia Sinica, Institute of Biomedical	
		REMARKS	Sciences, Rm 433, 128 Yen-Chun-Yuan Road SEC 2, Taipei 11529,	
		REMARKS	R.O.C Taiwan.	
		FEATURES	Location/Qualifiers	
			1. 274	
		source	/organism="Leishmania major"	
			/mol_type="genomic DNA"	
			/db_xref="taxon:5664"	
		CDS	/clone="LmV5-G1"	
			<1 .. >274	
			/codon_start=2	
			/evidence="not experimental"	
			/product="putative glycoprotein"	
			/protein_id="AAE0371.1"	
			/db_xref="GI:4731176"	
			/translation="EAEROKLKELQBEHEGEQEERARQRVAEEKAQKRAEKKAEEA EGELAARRRKGELELQLRHAEEKQRREMRKOREETQRKREI."	
		ORIGIN		
		Query Match	85.5%	Score 18.8; DB 3; Length 274;
		Best Local Similarity	90.9%	Pred. No. 6.1e+02;
		Matches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
		Qy	1 AAAGTGGAGTCCAGGAGCTGC 22	
		Db	167 AAGGGGAGTTGGAGGAGCTGC 188	
misc_feature		RESULT 13		
		LOCUS	LE1GP9692A	704 bp linear
		DEFINITION	Leishmania major glycoprotein 96-92 (GP 96-92) gene, partial	cds.
		ACCESSION	M63109	
		VERSION	M63109.1	GI:1593332
		SOURCE	Glucoprotein 96-92.	
		ORGANISM	Leishmania major	
		Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;		
		Leishmania.		
		REFERENCE	1 (bases 1 to 704)	
		AUTHORS	Handman,E., Barnett,L.D., Osborn,A.H., Godding,J.W. and Murray,P.J.	
		TITLE	Identification, characterisation and genomic cloning of a C-linked	
		JOURNAL	N-acetylglucosamine-containing cytoplasmic Leishmania glycoprotein	
		MEDLINE	Mol. Biochem. Parasitol. 62 (1), 61-72 (1993)	
		PUBLISHED	8114827	
		COMMENT	Original source text: Leishmania major DNA.	
		FEATURES	Location/Qualifiers	
			1 .. 704	
		source	/organism="Leishmania major"	
			/mol_type="unassigned DNA"	
Query Match	86.4%	Score 19; DB 1; Length 295150;		
Best Local Similarity	100.0%	Pred. No. 5.9e+02;		
Matches	19;	Conservative 0; Mismatches 0;	Indels 0; Gaps 0;	
Qy	3 GGTCGAGTCGAGGAGCTG 21			

/isolate="WR309"  
 /db\_xref="taxon:5664"  
 /clone\_lib="Lambda GT11"  
 1 . 704  
 /gene="GP 96-92"  
 <1 .>704  
 /gene="GP 96-92"  
 /codon\_start=3  
 /product="glycoprotein 96-92"  
 /protein\_id="AAI0179.1"  
 /db\_xref=GI:159333"  
 /translation="COAWLQDGWSWSPVSAGEEPEKORTEMVKQREPAQQKREIO  
 KQREEEIKRKAELAERQKKELEHEEEQEEARSRRVAAEKAQKKAKAEAE  
 DEIAATRQKOKGELEBLQRAEKEERQEEAKRQYRQVAAEKAQKKELEKAFAEDLAATRQK  
 GELEBLQAAQEEREEBF"

ORIGIN

Query Match	85.5%	Score 18.8;	DB 3;	Length 704;
Best Local Matches	90.9%	Pred. No. 6.2e+02;	Indels 2;	Gaps 0;
20; Conservative		Mismatches 2;		

Qy 1 AAGGTGGAGTTCGAGGAGCTGC 22  
 Db 339 AAGGGGGAGTTCGAGGAGCTGC 360

RESULT 14  
 AY051720 LOCUS AY051720 melanogaster 2315 bp mRNA linear INV 27-AUG-2001  
 DEFINITION Drosophila melanogaster LD24980 full length cDNA.  
 ACCESSION AY051720  
 VERSION AY051720.1  
 KEYWORDS GI:15291750  
 SOURCE FLI CDNA.  
 ORGANISM Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Ephemeroptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydriidae; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 2315)  
 AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,  
 Chapman,M., Chavez,C., Dorsett,V., Farren,D., Frise,E., George,R.,  
 Gonzalez,M., Guarini,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,  
 Nuno,J., Pacleb,J., Paraga,V.V., Park,S., Phouanenavong,S., Wan,K.,  
 Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.

TITLE Submitted (10 AUG-2001) Berkeley Drosophila Genome Project,  
 JOURNAL Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720 USA  
 COMMENT Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory  
 Berkeley, CA 94720 USA  
 This clone was sequenced as part of a high-throughput process to  
 sequence clones from Drosophila Gene Collection 1 (Rabin et al.,  
 Science 2000). The sequence has been subjected to integrity checks  
 for sequence accuracy, presence of a polyA tail and contiguity  
 within 100 kb in the genome. Thus we believe the sequence to  
 reflect accurately this particular cDNA clone. However, there are  
 artifacts associated with the generation of cDNA clones that may  
 have not been detected in our initial analyses such as internal  
 priming, priming from contaminating genomic DNA, retained introns  
 due to reverse transcription of unspliced precursor RNAs, and  
 reverse transcriptase errors that result in single base changes.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our Web site  
 (<http://fruitfly.berkeley.edu>) or send email to  
 cDNA@fruitfly.berkeley.edu.

FEATURES  
 Source  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /strain="Y; cm bw sp"

gene  
 gene "CG7846"  
 /note="Alignment with genomic scaffold AE003507"  
 /db\_xref="FLYBASE:FBgn0030877"  
 82 . 1905  
 /gene="CG7846"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product\_start=1  
 /product="LD24980P"  
 /protein\_id="AAK93144.1"  
 /db\_xref="GI:1591751"  
 /db\_xref="FLYBASE:FBgn0030877"  
 /translation="MORSPASSTSGRHAAPPPTGLPPOAQAPLEAPKLVITPGSSOHLIGRAADINPLILHAVAYRPGASDEPHDPPLPILLNNSNGLNLVEEBORLVL  
 VSRILQHQCVVDNKNRLRATVPQOLAHFRPSQAEKPSQGNQADEPWLDREARLKF  
 DDRILRLGAVIDARYNDIHEPIQORELNLYNEKGSSLQASMOHLERIWSTALEERLKIP  
 LRKGUTHCAVLVNDVTRRLRFETVLLRLERFCPLVQDSVASTTGGIGYGV  
 DAHILDEKEKECHINAASVCGAQPKHENURKHNQWLYRVTQDQTLQSLSTSAGYQPR  
 LLNTGTRKAVFTQAOVQDQYDCDPCDELYKETGRNGVRGGDLQLQSLSTSAGYQPR  
 PQLPVTDADDLIVDQDDETI SINCSQSLQGAGTAGQGMNSNGCYRGQGIVLVLQDQAI  
 QSNRLSSTETKRMFGSITLUVGSSAKLQGLAANLEQRISQVOSENVNLIKGMDAGM  
 VAWKGAAINSVLSARELMISQNNDWQRLGLRVLRERSPLW"

ORIGIN

Query Match	85.5%	Score 18.8;	DB 3;	Length 2315;
Best Local Matches	90.9%;	Pred. No. 6.4e+02;	Indels 0;	Gaps 0;
20; Conservative		Mismatches 0;		

Qy 1 AAGGTGGAGTTCGAGGAGCTGC 22  
 Db 358 ATGGTGGAGTTCGAGGAGCAGC 379

RESULT 15  
 AY059441 LOCUS AY059441 melanogaster 2905 bp mRNA linear INV 12-OCT-2001  
 DEFINITION Drosophila melanogaster GH08948 full length cDNA.  
 ACCESSION AY059441  
 VERSION AY059441.1  
 KEYWORDS FLI cDNA.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydriidae; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 2905)  
 AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,  
 Chapman,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,  
 Gonzalez,M., Guarini,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,  
 Nuno,J., Pacleb,J., Paraga,V.V., Park,S., Phouanenavong,S., Wan,K.,  
 Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.

TITLE Submitted (08-OCT-2001) Berkeley Drosophila Genome Project,  
 JOURNAL Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720 USA  
 COMMENT Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory  
 Berkeley, CA 94720 USA  
 This clone was sequenced as part of a high-throughput process to  
 sequence clones from Drosophila Gene Collection 1 (Rabin et al.,  
 Science 2000). The sequence has been subjected to integrity checks  
 for sequence accuracy, presence of a polyA tail and contiguity  
 within 100 kb in the genome. Thus we believe the sequence to  
 reflect accurately this particular cDNA clone. However, there are  
 artifacts associated with the generation of cDNA clones that may  
 have not been detected in our initial analyses such as internal  
 priming, priming from contaminating genomic DNA, retained introns  
 due to reverse transcription of unspliced precursor RNAs, and  
 reverse transcriptase errors that result in single base changes.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our Web site  
 (<http://fruitfly.berkeley.edu>) or send email to  
 cDNA@fruitfly.berkeley.edu.

FEATURES  
 Source  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /strain="Y; cm bw sp"

COMMENT Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory  
 Berkeley, CA 94720 USA  
 This clone was sequenced as part of a high-throughput process to  
 sequence clones from Drosophila Gene Collection 1 (Rabin et al.,  
 Science 2000). The sequence has been subjected to integrity checks  
 for sequence accuracy, presence of a polyA tail and contiguity  
 within 100 kb in the genome. Thus we believe the sequence to  
 reflect accurately this particular cDNA clone. However, there are  
 artifacts associated with the generation of cDNA clones that may  
 have not been detected in our initial analyses such as internal  
 priming, priming from contaminating genomic DNA, retained introns  
 due to reverse transcription of unspliced precursor RNAs and  
 reverse transcriptase errors that result in single base changes.

For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to [cDNA@fruitfly.berkeley.edu](mailto:cDNA@fruitfly.berkeley.edu).

## FEATURES

## source

## Location/Qualifiers

1.

.2905

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/map="161E1-61E2"

1.

.2905

/gene="Glat1"

/note="Alignment with genomic scaffold AE003470. gene does not completely overlap longest ORF"

/db\_xref="FLYBASE:FBgn0025593"

953..2755

/gene="Glat1"

/note="Longest ORF"

/codon\_start=1

/product="Glat1\_9948P"

/protein\_id="AALI347.1"

/db\_xref="GI:16076858"

/db\_xref="FLYBASE:FBgn0025593"

/translation="MSNLSEDYDYYVGGRRKTPPTTSTQQLTSPPVRLRPFEDAYR

GANINNGCRFYRHESFSTAPKRQHSSRDDESRUICHGEDEBATURQLLUDLQKOVSY

MSMNLSAKLDELORGDPHILETYVALECIPTQLOQUETYSVSEQSEVSEKRDMLWAKH

ELDTVQVKBEFELEYVDRLEBTHRKRLLEQSLTFLFTYLSFSAVLGMLOFGY

NTGVINAPERKNIENFKMDVYKDYGEDISEEFIQLYSVAVSI

ANRFGRKGGILLNNVGLAGACIMGETPKVSHSYEMULFGRETIGVNCGLNTSIVPMYI

SEIAPLNLRGGITVNRQLAUTVGLLISQVIGLEQLIGINBWPILLGLAICPAILOJL

LLPVCPESPSPVYLITKOWEEEARLARRRASGSVEDDIEMRAERAQOSESHTSTM

ELIICSPTRPPLIGIYVMOI-SQAFQSGINAVYYSTLUFMSSQTEBSAKFATIGCAGI

MVMTILVSPILMDRTGRFLHYGLGMFITSFIFITISFLIKASIVVSRIEBSCSC

RVMAPPNNAKMPASLGHLFVERPFDLHMPLKS"

## ORIGIN

Query Match 85.5%; Score 18.8; DB 3; Length 2905;

Best Local Similarity 90.9%; Pred. No. 6.5e+02; Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGTGGAGTCGAGGAGCTGC 22

Db 1457 AAGGAGGAGATCGAGGAGCTGC 1478

Search completed: July 18, 2004, 13:33:22  
Job time : 211.026 secs



GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:40:25 ; Search time 208 026 Seconds

(without alignments)  
4583.781 Million cell updates/sec

Title: US-10-051-769-6

Perfect score: 22

Sequence: 1 gtggaaaggccgcgtgtactcc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:

6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_dat:\*

7: gb\_ph:\*

8: gb\_dx1:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vt:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vt:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_Pln:\*

35: em\_htg\_rnd:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_htg\_hum:\*

40: em\_htg\_mus:\*

41: em\_htg\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
C 1	22	100.0	243412	2	AC127769	AC127769 Rattus no
C 2	20.4	92.7	2623	9	BC012186	BC012186 Homo sapi
C 3	20.4	92.7	2639	6	AX741159	AX741159 Sequence
C 4	20.4	92.7	2639	9	AK091746	AK091746 Homo sapi
C 5	20.4	92.7	3847	6	BD169701	BD169701 Human Gli
C 6	20.4	92.7	3856	9	BC028375	BC028375 Homo sapi
C 7	20.4	92.7	49052	9	AL359836	AL359836 Human Gli
C 8	20.4	92.7	68196	2	AC108407	AC108407 Mus muscu
C 9	20.4	92.7	120579	9	AC058887	AC058887 cirtb_173
C 10	20.4	92.7	221524	10	AC139040	AC139040 Mus muscu
C 11	18.8	85.5	960	6	AR070450	AR070450 Sequence
C 12	17.8	80.9	110000	2	LMFLICH36_16	Continuation (17 o
C 13	17.8	80.9	192759	2	AC141584	AC141584 Rattus no
C 14	17.4	79.1	1935	6	BD182190	BD182190 Polynucle
C 15	17.4	79.1	1935	6	BD188593	BD188593 Polynucle
C 16	17.4	79.1	8056	6	BD182188	BD182188 Polynucle
C 17	17.4	79.1	8056	6	BD188591	BD188591 Polynucle
C 18	17.4	79.1	9901	6	BD182263	BD182263 Polynucle
C 19	17.4	79.1	9901	6	BD188666	BD188666 Polynucle
C 20	17.2	78.2	1935	6	AX612540	AX612540 Sequence
C 21	17.2	78.2	3518	1	AY187875	AY187875 Escherich
C 22	17.2	78.2	6118	6	AF266215	AF266215 Cillachir
C 23	17.2	78.2	6118	6	E12430	E12430 DNA encodin
C 24	17.2	78.2	6118	6	AR065638	AR065638 Sequence
C 25	17.2	78.2	6118	6	AR065680	AR065680 Sequence
C 26	17.2	78.2	711	4	SSU81236	SSU81236 Sus scrofa
C 27	17.2	78.2	723	9	HSU81237	HSU81237 Human von W
C 28	17.2	78.2	750	6	AT8686	AT8686 Sequence 2
C 29	17.2	78.2	1076	4	AF061063	AF061063 Tupai glis
C 30	17.2	78.2	1076	4	TGII1624	TGII1624 Tupai glis
C 31	17.2	78.2	1155	4	AF427361	AF427361 Dasyurus k
C 32	17.2	78.2	1203	4	AF076480	AF076480 Chaetophr
C 33	17.2	78.2	1220	9	TBA410296	TBA410296 Tarsius b
C 34	17.2	78.2	1230	4	U97535	U97535 Manis sp. v
C 35	17.2	78.2	1231	4	MSU97535	MSU97535
C 36	17.2	78.2	1232	4	ZP1427365	ZP1427365 Megaptera
C 37	17.2	78.2	1233	4	AF226849	AF226849 Megaptera
C 38	17.2	78.2	1233	4	AF304109	AF304109 Megaptera
C 39	17.2	78.2	1233	4	DNO218158	DNO218158 Dasyurus n
C 40	17.2	78.2	1233	10	PTY251144	PTY251144 Petromus
C 41	17.2	78.2	1234	4	AF108834	AF108834 Physeter
C 42	17.2	78.2	1236	4	TMA427362	TMA427362 Topeute
C 43	17.2	78.2	1238	9	HLA410300	HLA410300 Hylobates
C 44	17.2	78.2	1242	4	ESE227364	ESE227364 Euphractu
C 45	17.2	78.2	1242	4	OPO4672	OPO4672 Octotona
					S78431	S78431 von Willebr

## ALIGNMENTS

RESULT 1	AC127769/c	LOCUS	243412 bp	DNA linear	HTG 19-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-157C16				WORKING DRAFT SEQUENCE, 3
UNORDERED PIECES					
AC127769	AC127769	AC127769	GI:25079525		
VERSION	AC127769.3				
KEYWORDS	HTG; HTGS; PHASEI; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Butheria; Rodentia; Muridae; Murinae;					
Rattus.					
REFERENCE	1 (bases 1 to 243412)				
AUTHORS	Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,				

Pred. No. is the number of results predicted by chance to have a

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anvallebechi,V., Ayogai,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnsted,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Bluyt,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davis,M.I., Davis,C., Davy-Carroll,L., De Andra,C., Dederich,D., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escoto,M., Eugene,A., Evans,C.A., Fallis,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Gunaratne,P., Haaland,W., Harnil,C., Hamilton,C., Hernandez,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.I., Hodgson,A., Hoguez,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Khan,Z., King,J., Koyar,C., Kowals,C., Kraft,C.L., Lebow,H., Lewis,J., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Louised H., Lizardo R.J., Lu,X., Ma,J., Maheshwari,M., Mahindarne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenan,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munida,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Okwuonu,G., Olaribunsgion,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,L., Pfamkooh,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-I., Puado,M., Quiroz,J., Rachlin,B., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steinle,M., Strong,R., Sutton,A., Svatsk,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,A., Tingey,S., Trejos,Z., Usmani,A., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Wilson,R., Wilesky,R., Woodin,R., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Smith,H.O., Weinstock,G. and Gibbs,R.A.

**TITLE** Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**AUTHORS** Rat Genome Sequencing Consortium.

**REFERENCE** Direct Submission

**JOURNAL** Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

**COMMENT** On Nov 19, 2002 this sequence version replaced gi:23269487. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence reads.

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GZSV  
Center clone name: CH230-157C16  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus Quality: 220496 bases at least Q40  
Consensus Quality: 222898 bases at least Q30  
Consensus Quality: 224623 bases at least Q20  
Estimated insert size: 224491; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 111919: contig of 111919 bp in length  
\* 111919: gap of unknown length  
\* 12020: 12020: contig of 6282 bp in length  
\* 12020: 12020: gap of unknown length  
\* 18302: 18302: gap of unknown length  
\* 18402: 18402: contig of 225011 bp in length.  
-----  
FEATURES  
source  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-157C16"  
/db\_xref="taxon:10116"  
1..1535  
misc\_feature  
/note="wgs contig"  
12020..13436  
misc\_feature  
/note="wgs contig"  
17209..18301  
misc\_feature  
/note="wgs contig"  
18402..19742  
misc\_feature  
/note="wgs contig"  
25075..27269  
misc\_feature  
/note="wgs contig"  
238339..242288  
misc\_feature  
/note="wgs contig"  
242339..243412  
/note="wgs contig"  
-----  
ORIGIN  
Query Match 100.0%; Score 22; DB 2; Length 243412;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2  
BC012186/c  
LOCUS BC012186  
DEFINITION Homo sapiens, clone IMAGE:4564853, mRNA  
VERSION BC012186  
SOURCE BC012186..1 GI:15082555  
ORGANISM Homo sapiens

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	FEATURES	Biotechnology (JP)
AUTHORS	1. .2639	SOURCE	Location/Qualifiers
TITLE			1. .2639
JOURNAL			/organism="Homo sapiens"
			/mol type="mRNA"
			/db_xref="taxon:9606"
REMARK			
COMMENT			
NIH-MGC Project URL:	<a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
Contact:	MGC help desk		
Tissue Procurement:	DCTD/DTP		
cDNA Library Preparation:	Rubin Laboratory		
DNA Sequencing by:	Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada		
info@bcgsc.ca			
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Ejell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsuka, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeary, Steven Ness, Pawan Pandoh, Anna-Liesa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasaia van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marras.			
Clone distribution:	MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>		
Series:	IRAL Plate: 29 Row: h Column: 17		
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer Frequency ORF analysis, GenomeScan Gene prediction.			
FEATURES			
SOURCE			
1. .2623			
/organism="Homo sapiens"			
/mol type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:4564853"			
/tissue type="Kidney, renal cell adenocarcinoma"			
/clone Lib="NIH MGC_14"			
/lab host="DH10 <sup>®</sup> -R"			
/note=Vector: pOTB7"			
ORIGIN			
Query Match	92.7%	Score 20.4; DB 9; Length 2623;	PAT 20-JUN-2003
Best Local Similarity	95.5%	Pred. No. 2.9e+02; Mismatches 1; Indels 0; Gaps 0;	
Matches	21;	Conservative	
VERSION	AX747159.1	GI:32131547	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1. .2639		
AUTHORS	Isozaki,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Irie,R., Tamechika,I., Seki,N., Yoshihikawa,T., Otsuka,M., Nagahara,K. and Masuho,Y.		
TITLE	Full-length cDNA sequences		
JOURNAL	Patent: EP 1308459-A 64 07-MAY-2003; Helix Research Institute (JP) ; Research Association for		
RESULT	3		
AX747159/c			
LOCUS	AX747159	2639 bp mRNA	linear
DEFINITION	Sequence 684 from Patent EP1308459.		
ACCESSION			
VERSION	AX747159		
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1. .2639		
AUTHORS	Isozaki,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Irie,R., Tamechika,I., Seki,N., Yoshihikawa,T., Otsuka,M., Nagahara,K. and Masuho,Y.		
TITLE	Full-length cDNA sequences		
JOURNAL	Patent: EP 1308459-A 64 07-MAY-2003; Helix Research Institute (JP) ; Research Association for		
RESULT	3		
AX747159/c			
LOCUS	AX747159	2639 bp mRNA	linear
DEFINITION	Sequence 684 from Patent EP1308459.		
ACCESSION			
VERSION	AX747159		
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1. .2639		
AUTHORS	Isozaki,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Irie,R., Tamechika,I., Seki,N., Yoshihikawa,T., Otsuka,M., Nagahara,K. and Masuho,Y.		
TITLE	Full-length cDNA sequences		
JOURNAL	Patent: EP 1308459-A 64 07-MAY-2003; Helix Research Institute (JP) ; Research Association for		
RESULT	4		
AK091746/c			
LOCUS	AK091746	2639 bp mRNA	linear
DEFINITION	Homo sapiens cDNA FLJ34427 Fis, clone HHDP02008816.		
ACCESSION	AK091746		
VERSION	AK091746.1	GI:21750192	
KEYWORDS	Oligo capping; Fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1. .2639		
AUTHORS	Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Imagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yanashita,H., Matsuo,Y., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi,Fuji,A., Oshima,A., Sugiyama,A., Kawakami,B., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K., and Isozaki,T.		
TITLE	NEO human cDNA sequencing project		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 2639), Isogai,T. and Yamamoto,T.		
JOURNAL	Direct Submission		
COMMENT	Submitted (04-JUL-2002) Takao Isozaki, BLJ Project (HRI Team); 2-6-7 kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3915, Fax: 81-438-52-3986		
	NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA Full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3' -end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.		
FEATURES	source		
	1. .2639		
	/organism="Homo sapiens"		
	/mol type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="HHDP02008816"		
	/cell type="dermal papilla cells (HDPC)"		
	/clone lib="HHDP02"		
	/note="cloning vector: pME18SFL3-primary culture, dermal papilla cells"		
ORIGIN			
Query Match	92.7%	Score 20.4; DB 9; Length 2639;	PAT 15-JUL-2002
Best Local Similarity	95.5%	Pred. No. 2.9e+02; Mismatches 0; Indels 0; Gaps 0;	
Matches	21;	Conservative	
VERSION	Qy	GI:32131547	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1. .2639		
AUTHORS	Isozaki,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Irie,R., Tamechika,I., Seki,N., Yoshihikawa,T., Otsuka,M., Nagahara,K. and Masuho,Y.		
TITLE	Full-length cDNA sequences		
JOURNAL	Patent: EP 1308459-A 64 07-MAY-2003; Helix Research Institute (JP) ; Research Association for		



ERIQKLEFMILDKLQNEIDQEELEHHNNSLVREKEETDTRKSSLSSAALAKSGERIQLAQT  
LIMMITYRAGIEDTIELETSLSLDQHKKISKYTDDEFLDNIEISOLIDSOPESSISDD  
1292 . .1546

/note="PDZ: Region: Domain present in PSD-95, Dlg, and  
ZO-1/2. Also called DHR (Dlg homologous region) or GGF  
(relatively well conserved tetrapeptide in these domains).  
Some PDZs have been shown to bind C-terminal polypeptides".  
/db\_xref="CDD;smart00228";  
2723 . .2851

/note="DAG PE-bind: Region: Phorbol esters/diacylglycerol  
binding domain (C1 domain). This domain is also known as  
the protein kinase C conserved region 1 (C1) domain".  
/db\_xref="CDD;pfam00130";

misc\_feature

Query Match 92.7%; Score 20.4; DB 9; Length 3856;  
Best Local Similarity 95.5%; Pred. No. 2.7e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

Query 1 GTGGAAGCCGCCGTTGACTCC 22  
Db 811 GTGGAAGCCGCCGTTGACTCC 790

RESULT 7

AL359836 AL359836 49052 bp DNA linear PRI 21-DEC-2001  
DEFINITION Human DNA sequence from clone RP11-389E6 on chromosome 10, complete  
sequence.

ACCESSION AL359836  
VERSION GI:17977720  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 49052)  
AUTHORS Smith, M.  
TITLE Direct Submission  
JOURNAL Submitted (21-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk. Clone requests: clonerquest@sanger.ac.uk  
COMMENT On Dec 23, 2001 this sequence version replaced gi:17384082. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPP; Information on the WORMPP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpp](http://www.sanger.ac.uk/Projects/C_elegans/wormpp) this sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrl0>. RP11-389E6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bapac/home.htm>

misc\_feature

Query Match 92.7%; Score 20.4; DB 9;  
Best Local Similarity 95.5%; Pred. No. 1.6e+01;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

Query Match 92.7%; Score 20.4; DB 9;  
Best Local Similarity 95.5%; Pred. No. 1.6e+01;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8

AC108407 AC108407 68196 bp DNA linear HTG 27-JAN-2002  
LOCUS Mus musculus clone RP24-422P10, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC108407  
ACCESSION AC108407.1  
VERSION GI:18377216  
KEYWORDS HTGS PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 68196)  
AUTHORS Birren, B.; Linton, L.; Musbbaum, C. and Lander, E.  
TITLE Unpublished  
JOURNAL  
REFERENCE 2 (bases 1 to 68196)  
AUTHORS Birren, B.; Linton, L.; Musbaum, C.; Landier, E.; Ali, A.; Allen, N.; Anderson, S.; Barna, N.; Bastian, V.; Boguslavskiy, L.; Boukhguler, B.; Brown, A.; Camarata, J.; LaRocque, K.; Chang, J.; Chazaro, B.; Choepel, Y.; Colangelo, M.; Collins, S.; Collymore, A.; Cook, A.; Cooke, P.; DeArellano, K.; Dewar, K.; Diaz, J.S.; Dodge, S.; Faro, S.; Ferreira, P.; FitzHugh, W.; Gage, D.; Galagan, J.; Gardyna, S.; Ginde, S.; Gord, S.; Goyette, M.; Graham, D.; Grandjean, J.; Grandjean, J.; Hagos, B.; Horcon, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kats, C.; LaRocque, K.; Lamazares, R.; Landers, T.; Lehoczy, J.; Levine, R.; Lin, G.; MacLean, C.; MacDonald, P.; Major, J.; Marquis, N.; Matthews, C.; McCarthy, M.; McEwan, P.; McErlean, K.; Meldrim, J.; Menes, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norby, C.; Norman, C.H.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, T.; Peterson, K.; Phunkhang, P.; Pierre, N.; Pollara, V.; Raymond, C.; Retta, R.; Riebach, M.; Riley, R.; Rose, C.; Rogov, P.; Roman, J.; Rosetti, M.; Roy, A.; Santos, R.; Schauer, S.; Schupbach, R.; Seaman, S.; Severy, P.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Strauss, N.; Subramanian, A.; Talaman, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Travis, N.; Trigilio, J.; Vassiliev, H.; Vialou, J.; Vo, A.; Wilson, P.; Wu, X.; Wyman, D.; Ye, W.J.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.

COMMENT Direct Submission Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center Center: Whitehead Institute / MIT Center for Genome Research Center code: WBR Web site: <http://www-seq.wi.mit.edu> Contact: sequence submission@genome.wi.mit.edu  
----- Project Information Center project name: Li8881

Center clone name: 422\_P\_10

\* NOTE: This record contains 85 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1: 714: contig of 714 bp in length

1: 814: gap of 100 bp  
 \* 815: 1520: contig of 706 bp in length  
 \* 1521: 1620: gap of 100 bp  
 \* 1621: 2323: contig of 703 bp in length  
 \* 2324: 2423: gap of 100 bp  
 \* 2424: 3134: contig of 711 bp in length  
 \* 3125: 3224: gap of 100 bp  
 \* 3225: 3937: contig of 703 bp in length  
 \* 3938: 4037: gap of 100 bp  
 \* 4038: 4739: contig of 702 bp in length  
 \* 4740: 4839: gap of 100 bp  
 \* 4840: 5545: contig of 706 bp in length  
 \* 5546: 5645: gap of 100 bp  
 \* 5646: 6352: contig of 707 bp in length  
 \* 6353: 6452: gap of 100 bp  
 \* 6453: 7172: contig of 720 bp in length  
 \* 7173: 7272: gap of 100 bp  
 \* 7273: 7996: contig of 724 bp in length  
 \* 7997: 8096: gap of 100 bp  
 \* 8097: 8811: contig of 718 bp in length  
 \* 8815: 8914: gap of 100 bp  
 \* 8915: 9629: contig of 715 bp in length  
 \* 9630: 9729: gap of 100 bp  
 \* 9730: 10411: contig of 688 bp in length  
 \* 10418: 10517: gap of 100 bp  
 \* 10518: 11223: contig of 706 bp in length  
 \* 11224: 11323: gap of 100 bp  
 \* 11324: 12014: contig of 691 bp in length  
 \* 12015: 12114: gap of 100 bp  
 \* 12115: 12816: contig of 702 bp in length  
 \* 12817: 12916: gap of 100 bp  
 \* 12917: 13616: contig of 700 bp in length  
 \* 13617: 13716: gap of 100 bp  
 \* 13717: 14422: contig of 706 bp in length  
 \* 14423: 14522: gap of 100 bp  
 \* 14523: 15234: contig of 712 bp in length  
 \* 15235: 15334: gap of 100 bp  
 \* 15335: 16043: contig of 709 bp in length  
 \* 16044: 16143: gap of 100 bp  
 \* 16144: 16859: contig of 716 bp in length  
 \* 16860: 16959: gap of 100 bp  
 \* 16960: 17669: contig of 701 bp in length  
 \* 17670: 20070: gap of 100 bp  
 \* 20071: 20171: contig of 704 bp in length  
 \* 20171: 20874: gap of 100 bp  
 \* 20875: 18461: contig of 709 bp in length  
 \* 18461: 18560: gap of 100 bp  
 \* 18561: 19269: contig of 710 bp in length  
 \* 19270: 19369: gap of 100 bp  
 \* 19370: 20070: contig of 701 bp in length  
 \* 20071: 20170: gap of 100 bp  
 \* 20171: 20874: contig of 704 bp in length  
 \* 20875: 21684: gap of 100 bp  
 \* 21685: 21784: contig of 710 bp in length  
 \* 21785: 22491: contig of 707 bp in length  
 \* 22492: 22591: gap of 100 bp  
 \* 22592: 23299: contig of 708 bp in length  
 \* 23299: 23399: gap of 100 bp  
 \* 23399: 24086: contig of 687 bp in length  
 \* 24086: 24186: gap of 100 bp

24187: \* 24901: contig of 715 bp in length  
 \* 24902: 25001: gap of 100 bp  
 \* 25002: 25711: contig of 713 bp in length  
 \* 25711: 25814: gap of 100 bp  
 \* 25815: 26528: contig of 714 bp in length  
 \* 26529: 26629: gap of 100 bp  
 \* 26629: 27358: contig of 730 bp in length  
 \* 27359: 27458: gap of 100 bp  
 \* 27459: 28159: contig of 701 bp in length  
 \* 28160: 28259: gap of 100 bp  
 \* 28260: 28966: contig of 707 bp in length  
 \* 28966: 29066: gap of 100 bp  
 \* 29066: 29765: contig of 699 bp in length  
 \* 29766: 29865: gap of 100 bp  
 \* 29866: 30577: contig of 712 bp in length  
 \* 30578: 30677: gap of 100 bp  
 \* 30678: 31381: contig of 704 bp in length  
 \* 31382: 31481: gap of 100 bp  
 \* 31482: 32192: contig of 711 bp in length  
 \* 32193: 32292: gap of 100 bp  
 \* 32293: 32974: contig of 682 bp in length  
 \* 32974: 33074: gap of 100 bp  
 \* 33075: 33760: contig of 686 bp in length  
 \* 33760: 34574: contig of 711 bp in length  
 \* 34574: 34674: gap of 100 bp  
 \* 34675: 35378: contig of 704 bp in length  
 \* 35379: 35478: gap of 100 bp  
 \* 35479: 36194: contig of 716 bp in length  
 \* 36194: 36294: gap of 100 bp  
 \* 36295: 36973: contig of 679 bp in length  
 \* 36974: 37073: gap of 100 bp  
 \* 37074: 37774: contig of 701 bp in length  
 \* 37775: 37875: gap of 100 bp  
 \* 37875: 38586: contig of 712 bp in length  
 \* 38586: 38686: gap of 100 bp  
 \* 38686: 39388: contig of 702 bp in length  
 \* 39388: 39488: gap of 100 bp  
 \* 39488: 40195: contig of 707 bp in length  
 \* 40195: 40295: gap of 100 bp  
 \* 40295: 41013: contig of 718 bp in length  
 \* 41014: 41113: gap of 100 bp  
 \* 41114: 41828: contig of 702 bp in length  
 \* 41828: 41928: gap of 100 bp  
 \* 41929: 42643: contig of 715 bp in length  
 \* 42644: 42743: gap of 100 bp  
 \* 42744: 43458: contig of 715 bp in length  
 \* 43458: 43558: gap of 100 bp  
 \* 43558: 44242: contig of 684 bp in length  
 \* 44243: 44342: gap of 100 bp  
 \* 44343: 45043: contig of 701 bp in length  
 \* 45044: 45144: gap of 100 bp  
 \* 45144: 45856: contig of 713 bp in length  
 \* 45857: 45956: gap of 100 bp  
 \* 45956: 46654: contig of 698 bp in length  
 \* 46655: 46754: gap of 100 bp  
 \* 46755: 47459: contig of 695 bp in length  
 \* 47459: 49986: gap of 100 bp  
 \* 49986: 50673: contig of 687 bp in length  
 \* 50673: 50773: gap of 100 bp  
 \* 50773: 51472: contig of 699 bp in length  
 \* 51472: 51572: gap of 100 bp  
 \* 51572: 52286: contig of 716 bp in length  
 \* 52286: 52389: gap of 100 bp  
 \* 52389: 53075: contig of 687 bp in length  
 \* 53075: 53175: gap of 100 bp  
 \* 53175: 53176: contig of 704 bp in length  
 \* 53176: 53879: contig of 704 bp in length

*	53880	53979:	gap of 100 bp			LOCUS	AC139040	221524 bp	DNA linear	ROD 16-JUL-2003	
*	53980	54675:	Contig of 696 bp in length			DEFINITION	Mus musculus chromosome 19, clone RP23-292H20, complete sequence.				
*	54676	54775:	gap of 100 bp			ACCESSION	AC139040				
*	54776	55477:	Contig of 702 bp in length			VERSION	AC139040.7				
*	55478	55577:	gap of 100 bp			KEYWORDS	HTG.				
Query Match		92.7%;	Score 20.4;	DB 2;	Length 68196;	ORGANISM	Mus musculus				
Best Local Similarity		95.5%;	Pred. No. 1.5e+02;				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Matches 21;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;			Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Mus.				
Qy	1	GTGGAAGGCCGCTGTACTCC 22				REFERENCE	1 (bases 1 to 221524)				
Db	17262	GTGAAACGGCCGTTGTACTCC 17283				AUTHORS	Birren, B.; Nusbaum, C. and Lander, E.				
						TITLE	Mus musculus chromosome 19, clone RP23-292H20				
						JOURNAL	Unpublished				
						REFERENCE	2 (bases 1 to 221524)				
						AUTHORS	Birren, B.; Nusbaum, C.; Lander, E.; Abouelleil, A.; Allen, N.; Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.				
RESULT 9						REFERENCE	3 (bases 1 to 221524)				
AC005887/C	AC005887	AC005887/citb_173_i_12,	complete sequence.	DNA	linear	AUTHORS	Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.				
LOCUS	AC005887	AC005887	citb_173_i_12,	complete sequence.	PRI 05-NOV-1999	TITLE	Submitted (23-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
DEFINITION						JOURNAL	Direct Submission				
VERSION		AC005887.3	GI:6249675			REFERENCE	4 (bases 1 to 221524)				
KEYWORDS						AUTHORS	Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.				
ORGANISM	Homo sapiens (human)					REFERENCE	Submitted (31-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE						AUTHORS	Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.				
AUTHORS	Smith,D.R.					JOURNAL	Direct Submission				
TITLE						REFERENCE	Submitted (29-OCT-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
JOURNAL						AUTHORS	Smith,D.R.				
REFERENCE						TITLE	Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
AUTHORS	Smith,D.R.					JOURNAL	Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
TITLE						REFERENCE	4 (bases 1 to 221524)				
JOURNAL						AUTHORS	Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.				
REFERENCE						JOURNAL	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
AUTHORS	Smith,D.R.					REFERENCE	Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
TITLE						AUTHORS	Smith,D.R.				
JOURNAL						TITLE	Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
REFERENCE						JOURNAL	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
AUTHORS	Smith,D.R.					REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
TITLE						AUTHORS	Smith,D.R.				
JOURNAL						TITLE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
REFERENCE						JOURNAL	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
AUTHORS	Smith,D.R.					REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
TITLE						AUTHORS	Smith,D.R.				
JOURNAL						TITLE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
COMMENT						JOURNAL	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
FEATURES						REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
source						AUTHORS	Smith,D.R.				
organism="Homo sapiens"						TITLE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
/mol type="genomic DNA"						JOURNAL	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
/db_xref="taxon:9606"						REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
/chromosome="10"						AUTHORS	Smith,D.R.				
/map="10q25"						TITLE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
/clone_lib="CIT987SK-117312"						JOURNAL	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
ORIGIN						REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA				
Query Match		92.7%;	Score 20.4;	DB 9;	Length 120578;	AUTHORS	Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.				
Best Local Similarity		95.5%;	Pred. No. 1.3e+02;	Mismatches 1;	Indels 0;	Gaps 0;	REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA			
Matches 21;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;			AUTHORS	Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.			
Qy	1	GTGGAAGGCCGCTGTACTCC 22					REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA			
Db	54219	GTGAAAGCCCCGTTGTACTCC 54198					AUTHORS	Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.			
RESULT 10							REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA			
AC139040/c							AUTHORS	Anderson, S.; Arachchi, H.M.; Barna, N.; Bastien, V.; Bloom, T.; Boguslavsky, L.; Bonkhgalter, B.; Camarata, J.; Chang, J.; Choepel, Y.; Collimore, A.; Cook, A.; Cooke, P.; Corum, B.; DeArellano, K.; Diaz, J.S.; Dodge, S.; Dooley, K.; Dorris, L.; Erickson, J.; Faro, S.; Ferreira, P.; Fitzgerald, M.; Gage, D.; Galagan, J.; Gardyna, S.; Graham, L.; Grand-Pierre, N.; Hafez, N.; Hagopian, D.; Hagos, B.; Hall, J.; Horton, L.; Huime, W.; Iliev, I.; Johnson, R.; Jones, C.; Kamat, A.; Karatas, A.; Kells, C.; Landers, T.; Levine, R.; Lindblad-Toh, K.; Liu, G.; Lui, A.; Mabbitt, R.; Maclean, C.; MacDonald, P.; Major, J.; Manning, J.; Matthews, C.; McCarthy, M.; Meldrim, J.; Maneus, L.; Mihova, T.; Mlenga, V.; Murphy, T.; Naylor, J.; Nguyen, C.; Nicol, R.; Norbu, C.; O'Connor, T.; O'Donnell, P.; O'Neill, D.; Oliver, J.; Peterson, K.; Phunkhang, P.; Pierre, N.; Rachupka, A.; Ramasamy, U.; Raymond, C.; Petta, R.; Rize, C.; Rogov, P.; Roman, J.; Schauer, S.; Schupback, R.; Seaman, S.; Severy, P.; Smith, C.; Spencer, B.; Stange-Thomann, N.; Stojanovic, N.; Stubbs, M.; Talama, J.; Tesfaye, S.; Theodore, J.; Topham, K.; Travers, M.; Vassiliev, H.; Venkataraman, V.S.; Viel, R.; Vo, A.; Wilson, B.; Wu, X.; Wyman, D.; Young, G.; Zainoun, J.; Zembek, L.; Zimmer, A. and Zody, M.			

Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
 Oliver,J., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Riese,C., Rogov,P.,  
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severny,P., Smith,C.,  
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Tesfaye,S., Theodore,J., Tophan,K., Travers,M.,  
 Vassiliev,H., Venkataraman,Y.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wymam,D., Young,G., Zainoun,J., Zembek,I., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (16-JUL-2003) Street, Cambridge, MA 02141, USA  
 Research, On Jul 16, 2003 this sequence version replaced gi:31249817.  
 All repeats were identified using RepeatMasker:  
 Smits,A.F.A. & Green,P. (1996-1997)  
 http://Ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L20063  
 Center clone name: 292\_H\_20

---

**FEATURES**

source	Location/Qualifiers
1..221524	/clone lib="RPCI-23 Female Mouse BAC"
	/note="WGS end_extension
	clone_end_SP6"
misc_feature	
unsure	
repeat_region	/note="<30 qual_SNGL region"
repeat_region	/clone end_SP6"
repeat_region	/note="<30 qual_SNGL region"
repeat_region	/clone lib="RPCI-23 Female Mouse BAC"
repeat_region	/note="WGS end_extension
repeat_region	clone_end_SP6"
repeat_region	/note="<30 qual_SNGL region"
repeat_region	/rpt_family="MUR1H1"
repeat_region	/rpt_complement(1981..2123)
repeat_region	/rpt_family="MER117"
repeat_region	/rpt_complement(2143..2511)
repeat_region	/rpt_complement(2897..2955)
repeat_region	/rpt_family="CA"
repeat_region	/rpt_complement(3063..3114)
repeat_region	/rpt_complement(3141..3217)
repeat_region	/rpt_complement(3141..3217)
repeat_region	/rpt_complement(3235..3402)
repeat_region	/rpt_complement(3447..3584)
repeat_region	/rpt_complement(3446..3690)
repeat_region	/rpt_complement(3724..3752)
repeat_region	/rpt_complement(4335..4351)
repeat_region	/rpt_complement(4352..4408)
repeat_region	/rpt_complement(4409..4625)
repeat_region	/rpt_complement(4701..4799)
repeat_region	/rpt_complement(5000..5112)
repeat_region	/rpt_complement(5268..5451)
repeat_region	/rpt_complement(5782..5810)
repeat_region	/rpt_complement(5810..5841)

repeat\_region /note="clone boundary clone\_end\_SP6"  
 site:Ecoli "

repeat\_region 7196..7246  
 /rpt\_complement(736..7656)  
 /rpt\_family="ORR1C"

repeat\_region 8336..8384  
 /rpt\_complement(913..9122)  
 /rpt\_family="NIR"

repeat\_region 8968..9042  
 /rpt\_complement(913..9332)  
 /rpt\_family="B\_Mm2"

repeat\_region 1047..7052  
 /rpt\_complement(984..10055)  
 /rpt\_family="B3"

repeat\_region 11638..11844  
 /rpt\_complement(1184..12494)  
 /rpt\_family="B3"

repeat\_region 12218..12494  
 /rpt\_complement(1373..139993)  
 /rpt\_family="B4"

repeat\_region 14922..14986  
 /rpt\_complement(1570..15731)  
 /rpt\_family="ID\_B1"

repeat\_region 15105..15218  
 /rpt\_complement(1570..15731)  
 /rpt\_family="L1M4"

repeat\_region 15345..15509  
 /rpt\_complement(1570..15731)  
 /rpt\_family="ID\_B1"

repeat\_region 15732..15849  
 /rpt\_complement(1570..15731)  
 /rpt\_family="ORR1D"

repeat\_region 15850..15874  
 /rpt\_complement(1570..15731)  
 /rpt\_family="TGGGGG"

repeat\_region 15875..16100  
 /rpt\_complement(1570..15731)  
 /rpt\_family="ORR1D"

repeat\_region 16101..16251  
 /rpt\_complement(1610..16251)  
 /rpt\_family="ID\_B1"

repeat\_region 17419..17513  
 /rpt\_complement(1610..16251)  
 /rpt\_family="CA"

repeat\_region 17656..17653  
 /rpt\_complement(1610..16251)  
 /rpt\_family="CCCA"

repeat\_region 18312..18451  
 /rpt\_complement(1610..16251)  
 /rpt\_family="ID\_B1"

repeat\_region 18451..18730  
 /rpt\_complement(1610..16251)  
 /rpt\_family="RMR10A"

Query Match 92..7%; Score: 20..4; DB 10;  
 Best Local Similarity 95..%; Pred. No. 1..1e+02;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAGCGCGCTGTGTACTCC 22  
 Db 43076 GTGGAAACCGCGTGTACTCC 43055

RESULT 11  
 AR070450/c AR070450  
 LOCUS DEFINITION Sequence 1 from patent US 5900476.  
 ACCESSION AR070450  
 VERSION GI:7221338  
 SOURCE Unknown.  
 ORGANISM Unknown.

Unclassified.			
1 (bases 1 to 960)			
AUTHORS Ruggeri, Z.M. and Ware, J.L.			
TITLE Therapeutic domains of van Willebrand factor			
PATENT: US 5900476-A 1. 04-MAY-1999;			
FEATURES 1. .960			
JOURNALS /organism="unknown"			
source /mol_type="unassigned DNA"			
ORIGIN			
Query Match 85.5%; Score 18.8; DB 6; Length 960;			
Best Local Similarity 90.9%; Pred. No. 1.7e+03; Mismatches 0; Indels 0; Gaps 0;			
Matches 20; Conservative 0; Qualifiers			
Qy 1 GTGGAAGCCGCCGTTGTACTCC 22			
Db 399 GTGGAAGCCGCCGTTGTACTCC 378			
RESULT 12			
LMLCHR36_16/c			
COMMENT Sequence split into 36 fragments LOCUS LMLCHR36 Accession AL499624			
Fragment Name Begin End			
LMLCHR36_00 1 110000			
LMLCHR36_01 100001 210000			
LMLCHR36_02 200001 310000			
LMLCHR36_03 300001 410000			
LMLCHR36_04 400001 510000			
LMLCHR36_05 500001 610000			
LMLCHR36_06 600001 710000			
LMLCHR36_07 700001 810000			
LMLCHR36_08 800001 910000			
LMLCHR36_09 900001 1010000			
LMLCHR36_10 1000001 1110000			
LMLCHR36_11 1100001 1210000			
LMLCHR36_12 1200001 1310000			
LMLCHR36_13 1300001 1410000			
LMLCHR36_14 1400001 1510000			
LMLCHR36_15 1500001 1610000			
LMLCHR36_16 1600001 1710000			
LMLCHR36_17 1700001 1810000			
LMLCHR36_18 1800001 1910000			
LMLCHR36_19 1900001 2010000			
LMLCHR36_20 2000001 2110000			
LMLCHR36_21 2100001 2210000			
LMLCHR36_22 2200001 2310000			
LMLCHR36_23 2300001 2410000			
LMLCHR36_24 2400001 2510000			
LMLCHR36_25 2500001 2610000			
LMLCHR36_26 2600001 2710000			
LMLCHR36_27 2700001 2810000			
LMLCHR36_28 2800001 2910000			
LMLCHR36_29 2900001 3010000			
LMLCHR36_30 3000001 3110000			
LMLCHR36_31 3100001 3210000			
LMLCHR36_32 3200001 3310000			
LMLCHR36_33 3300001 3410000			
LMLCHR36_34 3400001 3510000			
LMLCHR36_35 3500001 3529832			
Continuation (17 of 36) of LMLCHR36 from base 1600001 (AL499624 Leishmania major chromo			
Query Match 80.9%; Score 17.8; DB 2; Length 110000;			
Best Local Similarity 90.5%; Pred. No. 1.7e-03; Mismatches 0; Indels 0; Gaps 0;			
Matches 19; Conservative 0; Qualifiers			
Qy 2 TGGAAAGCCGCCGTTGTACTCC 22			

of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Mar 19, 2003 this sequence version replaced gi:29029160.  
 COMMENT  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: KEEH  
 Center clone name: CH230-247A23  
 Summary Statistics  
 Sequencing vector: Plasmid,  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap; version 0\_990329  
 Consensus quality: 164290 bases at least Q40  
 Consensus quality: 173116 bases at least Q30  
 Consensus quality: 179982 bases at least Q20  
 Estimated insert size: 176498; sum-of-contigs estimation  
 Quality coverage: 2x in Q20 bases; sum-of-contigs estimation  
 NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Gerbank/draft\\_data.htm](http://www.hgsc.bcm.tmc.edu/docs/Gerbank/draft_data.htm))  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 72 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1312: contig of 1312 bp in length  
 \* 1412: gap of unknown length  
 \* 2496: contig of 1084 bp in length  
 \* 596: gap of unknown length  
 \* 2597: contig of 1365 bp in length  
 \* 3961: gap of unknown length  
 \* 4061: gap of 1303 bp in length  
 \* 5364: gap of unknown length  
 \* 5365: contig of 1249 bp in length  
 \* 6713: gap of unknown length  
 \* 6714: gap of 1141 bp in length  
 \* 6814: gap of unknown length  
 \* 7954: contig of 1354 bp in length  
 \* 7955: gap of unknown length  
 \* 8055: contig of 1348 bp in length  
 \* 9409: gap of unknown length  
 \* 9509: contig of 1045 bp in length  
 \* 10856: gap of unknown length  
 \* 10956: gap of unknown length  
 \* 10957: contig of 1286 bp in length  
 \* 12242: gap of unknown length  
 \* 12343: contig of 1351 bp in length  
 \* 13387: gap of unknown length  
 \* 13488: gap of 1150 bp in length  
 \* 14638: gap of unknown length  
 \* 14637: contig of 14737 bp in length  
 \* 14738: gap of 1158 bp in length  
 \* 15895: contig of 1464 bp in length  
 \* 15896: gap of unknown length  
 \* 15996: contig of 1371 bp in length  
 \* 17347: gap of unknown length  
 \* 17447: gap of 1464 bp in length  
 \* 18910: contig of 1833 bp in length  
 \* 18911: gap of unknown length  
 \* 19011: contig of 1464 bp in length  
 \* 20474: gap of unknown length  
 \* 20574: gap of unknown length  
 \* 2075: contig of 2088 bp in length  
 \* 21945: gap of unknown length  
 \* 22045: gap of 1113 bp in length  
 \* 22046: contig of 1661 bp in length  
 \* 23458: gap of unknown length  
 \* 23659: gap of unknown length  
 \* 23759: contig of 1425 bp in length  
 \* 25591: gap of unknown length  
 \* 25592: gap of unknown length  
 \* 25691: gap of unknown length  
 \* 25692: contig of 2088 bp in length  
 \* 25692: contig of 1371 bp in length  
 \* 27780: gap of unknown length  
 \* 27880: contig of 1661 bp in length  
 \* 29541: gap of unknown length  
 \* 29641: contig of 1425 bp in length  
 \* 31065: gap of unknown length  
 \* 31165: gap of unknown length

KEYWORDS WO 02095028-A/3.  
 SOURCE Lactobacillus brevis  
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 Lactobacillus  
 1 (bases 1 to 1935)  
 REFERENCE Fujii,T.  
 AUTHORS Polynucleotide probe and primer for detecting beer-clouding lactic acid bacterium and method of detecting beer-clouding lactic acid  
 JOURNAL Patent: WO 02095028-A 3 28-NOV-2002;  
 KIRIN BREWERY CO LTD TOSHIRO FUJII  
 OS Lactobacillus brevis  
 PN WO 02095028-A/3  
 PD 28-NOV-2002  
 PP 23-MAY-2002 WO 200201P005022  
 PR 23-MAY-2001 JP 01P 154085  
 PI TOSHIRO FUJII  
 PC C12N15/11,C12N1/15, C12N1/19,C12N1/21,C12N5/10, C07K14/335, PC  
 C07KL6/12,  
 C12P21/02,C12Q1/04,C12Q1/68  
 CC Polynucleotide probe and primer for detecting beer-clouding lactic acid  
 CC lactic acid  
 CC bacterium and method of detecting beer-clouding lactic acid  
 FH Key  
 FT CDS Location/Qualifiers  
 1. (1932).  
 FEATURES source  
 1. 1935  
 LOCATION/QUALIFIERS  
 /organism="Lactobacillus brevis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1580"  
 ORIGIN

Query Match 79.1%; Score 17.4; DB 6; Length 1935;  
 Best Local Similarity 94.7%; Pred. No. 5.7e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGGCCGCGTTGTAC 19  
 Db 1146 GCGAAAGGCCGCGTTGTAC 1128

Search completed: July 18, 2004, 13:33:25  
 Job time : 211.026 secs

---

RESULT 15  
 LOCUS BD188593/c  
 DEFINITION BD188593 1935 bp DNA linear PAT 17-JUL-2003  
 Polynucleotide probe and primer for detecting beer-clouding lactic acid bacterium and method of detecting beer-clouding lactic acid bacterium.  
 ACCESSION BD188593  
 KEYWORDS JP 2003000251-A/3.  
 SOURCE Lactobacillus brevis  
 ORGANISM Lactobacillus brevis  
 Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 Lactobacillus  
 1 (bases 1 to 1935)  
 REFERENCE Fujii,T.  
 AUTHORS Polynucleotide probe and primer for detecting beer-clouding lactic acid bacterium and method of detecting beer-clouding lactic acid  
 JOURNAL Patent: JP 2003000251-A 3 07-JAN-2003;  
 KIRIN BREWERY CO LTD  
 OS Lactobacillus brevis  
 PN JP 2003000251-A/3  
 PD 07-JAN-2003  
 PP 23-MAY-2001 JP 2001154085  
 PI TOSHIRO FUJII  
 PC C12N15/09,C07K14/335, C07K16/12, C12N1/15,C12N1/19,C12N1/21, PC  
 C12N5/10,  
 C12P21/02,C12Q1/68,G01N33/14,G01N33/53,G01N33/566,G01N33/569//  
 PC C12Q1/68,C12R1/24,C12N15/00,C12N5/00  
 CC Polynucleotide probe and primer for detecting beer-clouding lactic acid



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:31:29 ; Search time 36.7148 Seconds

(without alignments)  
2545.580 Million cell updates/sec

Title: US-10-051-769-5  
Perfect score: 22

Sequence: 1 aaggtrgagttcgaggagctgc 22

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:

6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: geneseqn1980s:\*
- 2: geneseqn1990s:\*
- 3: geneseqn2000s:\*
- 4: geneseqn2001as:\*
- 5: geneseqn2001bs:\*
- 6: geneseqn2002s:\*
- 7: geneseqn2003as:\*
- 8: geneseqn2003bs:\*
- 9: geneseqn2003cs:\*
- 10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	7 ABX95682	Abx95682 Human G11
2	22	100.0	22	9 AAD61562	Aad61562 Human G11
3	22	100.0	261	7 ABX95679	Abx95679 Human G11
4	22	100.0	261	9 AAD61560	Aad61560 Human G11
5	22	100.0	263	7 ABX95681	Abx95681 Rat Glute
6	22	100.0	263	9 AAD61561	Aad61561 Human G11
7	22	100.0	673	7 ABX95680	Abx95680 Human G11
8	22	100.0	960	7 ABX95678	Abx95678 Human Gen
9	22	100.0	960	9 AAD61559	Aad61559 Human G11
10	22	100.0	1092	9 ADE07789	Ade07789 Novel cod
11	22	100.0	2639	9 ADB62530	Adb62530 Human CDN
12	22	100.0	3832	9 AAD61564	Aad61564 Human G11
13	22	100.0	4450	4 AAKS1608	Aak51608 Human P01
14	22	100.0	4470	4 ABA09174	Aba09174 Human sec
15	22	100.0	4470	4 AAKS2592	Aak52592 Human P01
16	22	100.0	4470	9 ADE03891	Ade03891 Novel DNA
17	22	100.0	4801	9 ADD49052	Add49052 Human NOV
c 18	19.4	88.2	1732	6 ABQ16934	Abq16934 Oligomeric
c 19	18.8	95.5	2249	4 ABL13453	Ab113453 Drosophili
c 20	18.8	85.5	3582	4 ABL1714	Ab11714 Drosophili
c 21	18.8	85.5	4382	4 ABL13452	Ab113452 Drosophili
c 22	18.8	85.5	5973	6 ABS78752	Ab878752 DNA encod

#### ALIGNMENTS

24	18.8	85.5	6202	7 ABX70893	Novel hum
25	18.8	85.5	7584	7 ABZ34402	Human sec
26	18.8	85.5	26781	4 ABL01850	Drosophil
27	18.8	85.5	35498	4 ABL20316	Drosophil
28	18.4	83.6	2459	4 ABL12495	Drosophil
29	18.4	83.6	6942	4 ABL12494	Drosophil
30	17.8	80.9	190	4 AAI27112	Probe #17
31	17.8	80.9	190	4 ABA15387	Human fce
32	17.8	80.9	190	4 AAI15954	Probe #24
33	17.8	80.9	190	4 ABA0026	Probe #18
34	17.8	80.9	190	4 AAK50017	Human bon
35	17.8	80.9	190	4 AAK3948	Human bra
36:	17.8	80.9	190	4 ABS19663	Human liv
37:	17.8	80.9	190	6 ABS2510	Human gen
c 38	17.8	80.9	327	8 ACL16432	DNA clone
c 39	17.8	80.9	336	8 ACL16445	DNA clone
c 40	17.8	80.9	366	8 ACL16430	DNA clone
c 41	17.8	80.9	385	8 ACL16414	DNA clone
c 42	17.8	80.9	426	4 ABL27111	Drosophil
c 43	17.8	80.9	450	8 ACL16416	DNA clone
c 44	17.8	80.9	450	8 ACL16447	DNA clone
c 45	17.8	80.9	460	8 ACL16443	DNA clone

#### RESULT 1

ABX95682	ID ABX95682 standard; DNA; 22 BP.
XX	AC ABX95682;
AC	XX
XX	DT 30-JUN-2003 (first entry)
XX	DE Human GLTEN PCR primer #1.
XX	XX Human; #6; PCR; chromosome 10q25; glioblastoma multiforme; brain cancer; KW GLTEN; GMB; primer.
XX	OS Homo sapiens.
XX	PN US200304811-A1.
XX	PD 06-MAR-2003.
XX	PP 20-OCT-2001; 2001US-00051769.
PR	XX 20-OCT-2000; 2000US-0242160B.
PR	XX 06-MAR-2003.
PA (UYNE) UNIV NEW JERSEY MEDICINE & DENTISTRY.	XX
PA	XX Mckinnon RD;
XX	DR WPI; 2003-418995/39.
XX	PT New isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLTEN, useful as a molecular marker for tumor cell identification and classification.
PS	Claim 5; Page 2; Lipp; English.
XX	The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLTEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridizes under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme, designated as GLTEN, useful as a molecular marker for tumor cell identification and classification.
CC	The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLTEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridizes under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control

CC sample to assess the level of mRNA in a control sample, where the  
 CC presence of increased levels of mRNA expression in the sample in an  
 CC amount higher than the control sample indicates risk for progression into  
 CC Glioblastoma Multiforme, and a kit for use in detecting whether a  
 CC patient is at risk for progression into Glioblastoma Multiforme  
 CC (comprising the probe or primers). The probe or the kit is useful for  
 CC detecting whether a patient is at risk for progression into Glioblastoma  
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker  
 CC for tumour cell identification and classification and for diagnosing or  
 CC identifying candidates at risk for progression into a malignant phenotype  
 CC especially in brain cancer therapy, treatment and diagnosis. The present  
 CC sequence is a human GLITEN gene PCR primer  
 XX

SQ Sequence 22 BP; 5 A; 3 C; 10 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 7; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGGTGGAGTTCGAGGAGCTGC 22  
 1 AACGTGGAGTTCGAGGAGCTGC 22

RESULT 2

ID ADD61562  
 ID ADD61562 standard; DNA; 22 BP.  
 XX

AC ADD61562;

XX DT 15-JAN-2004 (first entry)

DE Human GLITEN cDNA amplifying PCR primer #1.

XX Human; Glioblastoma multiforme; GBM; GLITEN; brain cancer; diagnosis;

KW Gene therapy; PCR; primer; 5S.

XX Homo sapiens.

OS US2003108915-A1.

PN XX

PD 12-JUN-2003.

XX PR 20-AUG-2002; 2002US-002244624.

XX PR 20-OCT-2000; 2000US-0242166P.

XX PR 20-OCT-2001; 2001US-00051799.

PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX PT McKinnon RD;

XX DR 2003-810848/76.

XX PT Novel Glioblastoma multiforme associated protein GLITEN useful for

CC treating Glioblastoma multiforme and diagnosing brain cancer.

XX PS Claim 16; Page 5; Opp; English.

CC The present invention provides novel Glioblastoma multiforme (GBM)

CC associated protein GLITEN useful for treating glioblastoma multiforme and

CC diagnosing brain cancer. The invention is useful in brain cancer therapy,

CC treatment and diagnosis. The invention is also useful in gene therapy.

CC The present sequence is human GLITEN cDNA amplifying PCR primer

CC Sequence 22 BP; 5 A; 3 C; 10 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 7; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGGTGGAGTTCGAGGAGCTGC 22

Query Match 100.0%; Score 22; DB 9; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGGTGGAGTTCGAGGAGCTGC 22

Query Match 100.0%; Score 22; DB 7; Length 261;

Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 AACGTGGAGTTCGAGGAGCTGC 22

Oy

Db 1 AAGGTGGAGTTCGAGGAGCTGC 22

RESULT 3

ID ABX95679 standard; cDNA; 261 BP.  
 XX ABX95679;

AC AC;

XX DT 30-JUN-2003 (first entry)

XX DE Human GLITEN expressed sequence tag.  
 XX KW Human; ss; EST; chromosome 10q25; glioblastoma multiforme; brain cancer;  
 XX GLITEN; GMB; expressed sequence tag.

OS Homo sapiens.

PN US2003044811-A1.

XX PD 06-MAR-2003.

XX PR 20-OCT-2001; 2001US-00051769.

XX PR 20-OCT-2000; 2000US-0242160P.

XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX PI McKinnon RD;

XX DR 2003-418995/39.

XX PT New isolated nucleic acid representing a gene product associated with

PT Glioblastoma Multiforme, designated as GLITEN, useful as a molecular

PT marker for tumor cell identification and classification.

XX PS Claim 1; Page 7; 11pp; English.

XX The invention relates to an isolated nucleic acid representing a gene

CC product associated with Glioblastoma Multiforme, designated as GLITEN,

CC comprising the human EST (expressed sequence tag) appearing as ABX95678,

CC or a sequence that hybridises under stringent conditions to the EST, or

CC its complement. Also included are a probe for use in identifying a

CC patient at risk for progression into the malignant phenotype, comprising

CC the nucleic acids detailed above, detecting whether a patient is at risk

CC for progression into Glioblastoma Multiforme (comprising: (a) providing a

CC sample from a patient; (b) adding the probe to the sample or performing

CC PCR analysis using the primers appearing as ABX95612 and ABX95683; (c)

CC analysing levels of mRNA bound with the probe; and (d) treating a control

CC sample to assess the level of mRNA in a control sample, where the

CC presence of increased levels of mRNA expression in the sample in an

CC amount higher than the control sample indicates risk for progression into

CC Glioblastoma Multiforme), and a kit for use in detecting whether a

CC patient is at risk for progression into Glioblastoma Multiforme

CC (comprising the probe or primers). The probe or the kit is useful for

CC detecting whether a patient is at risk for progression into Glioblastoma

CC Multiforme (a type of brain cancer). The EST useful as molecular marker

CC for tumour cell identification and classification and for diagnosing or

CC identifying candidates at risk for progression into a malignant phenotype

CC especially in brain cancer therapy, treatment and diagnosis. The present

CC sequence is the human GLITEN expressed sequence tag

XX Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

DR WPI; 2003-810848/76.

XX PT Novel Glioblastoma multiforme associated protein GLITEN useful for

CC treating Glioblastoma multiforme and diagnosing brain cancer.

XX PS Claim 16; Page 5; Opp; English.

CC The present invention provides novel Glioblastoma multiforme (GBM)

CC associated protein GLITEN useful for treating glioblastoma multiforme and

CC diagnosing brain cancer. The invention is useful in brain cancer therapy,

CC treatment and diagnosis. The invention is also useful in gene therapy.

CC The present sequence is human GLITEN cDNA amplifying PCR primer

CC Sequence 22 BP; 5 A; 3 C; 10 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 7; Length 261;

Best Local Similarity 100.0%; Pred. No. 5.6;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 AACGTGGAGTTCGAGGAGCTGC 22

Oy

**RESULT 4**  
 AAD61560 standard; cDNA; 261 BP.  
 XX ID AAD61560;  
 AC AAD61560;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Human Gliten EST clone cDNA #1.  
 XX KW Human; glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;  
 KW gene therapy; expressed sequence tag; ss.  
 XX OS Homo sapiens.  
 XX PN US2003108915-A1.  
 XX PD 12-JUN-2003.  
 XX PT 20-AUG-2002; 2002US-00224624.  
 XX PR 20-OCT-2000; 2000US-0242160P.  
 PR 20-OCT-2001; 2001US-00051769.  
 XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 XX PI McKinnon RD;  
 XX DR WPI; 2003-418995/39.  
 AX New isolated nucleic acid representing a gene product associated with  
 PT Glioblastoma Multiforme, designated as Gliten useful as a molecular  
 marker for tumor cell identification and classification.  
 XX Example 2; Page 4; 11pp; English.

**RESULT 5**  
 ABX95581 standard; cDNA; 263 BP.  
 XX ID ABX95581;  
 AC ABX95581;  
 XX DT 30-JUN-2003 (First entry)  
 XX DB Rat Gliten expressed sequence tag.  
 XX KW Rat; ss; EST; glioblastoma multiforme; brain cancer; Gliten; GMB;  
 expressed sequence tag.  
 XX OS Rattus Norvegicus.  
 XX PN US2003044811-A1.  
 XX PD 06-MAR-2003.  
 XX PR 20-OCT-2001; 2001US-00051769.  
 PF 20-OCT-2000; 2000US-0242160P.  
 PR 20-OCT-2001; 2001US-00051769.

---

**XX PR 20-OCT-2000; 2000US-0242160P.**  
 XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 XX XX  
 XX PI McKinnon RD;  
 XX XX  
 XX DR WPI; 2003-418995/39.  
 AX New isolated nucleic acid representing a gene product associated with  
 PT Glioblastoma Multiforme, designated as Gliten useful as a molecular  
 marker for tumor cell identification and classification.  
 XX Example 2; Page 4; 11pp; English.  
 XX AX  
 CC The invention relates to an isolated nucleic acid representing a gene  
 CC product associated with Glioblastoma Multiforme, designated as Gliten,  
 CC comprising the human EST (expressed sequence tag) appearing as ABX95670,  
 CC or a sequence that hybridises under stringent conditions to the EST, or  
 CC its complement. Also included are a probe for use in identifying a  
 CC patient at risk for progression into the malignant phenotype, comprising  
 CC the nucleic acids detailed above, detecting whether a patient is at risk  
 CC for progression into Glioblastoma Multiforme (comprising: (a) providing a  
 CC sample from a patient; (b) adding the probe to the sample; (c) performing  
 CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)  
 CC analysing levels of mRNA bound with the probe; and (d) treating a control  
 CC sample to assess the level of mRNA in a control sample, where the  
 CC presence of increased levels of mRNA expression in the sample in an  
 CC amount higher than the control sample indicates risk for progression into  
 CC Glioblastoma Multiforme, and a kit for use in detecting whether a  
 CC patient is at risk for progression into Glioblastoma Multiforme  
 (comprising the probe or the kit is useful for  
 CC detecting whether a patient is at risk for progression into Glioblastoma  
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker  
 CC for tumour cell identification and classification and for diagnosing or  
 CC identifying candidates at risk for progression into a malignant phenotype  
 CC especially in brain cancer therapy, treatment and diagnosis. The present  
 CC sequence is a Rat Gliten expressed sequence tag  
 XX SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;  
 XX Query Match 100.0%; Score 22; DB 7; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AGGTGGAGTTCCAGGAGCTGC 22  
 Db 5 AGGTGGAGTTCCAGGAGCTGC 26  
 RESULT 6  
 AAD61561 standard; cDNA; 263 BP.  
 XX ID AAD61561;  
 AC AAD61561;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Human Gliten EST clone cDNA #2.  
 XX KW Human; Glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;  
 KW gene therapy; expressed sequence tag; ss.  
 XX OS Homo sapiens.  
 XX PN US2003108915-A1.  
 XX PD 12-JUN-2003.  
 XX PF 20-AUG-2002; 2002US-00224624.  
 XX PR 20-OCT-2000; 2000US-0242160P.  
 PR 20-OCT-2001; 2001US-00051769.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 PA  
 XX Mckinnon RD;  
 PI  
 XX WPI; 2003-810848/76.  
 PT Novel glioblastoma multiforme associated protein Gliten useful for treating glioblastoma multiforme and diagnosing brain cancer.  
 PS Example 2; Page 8; OPP; English.  
 XX The present invention provides novel glioblastoma multiforme (GBM) associated protein Gliten useful for treating glioblastoma multiforme and diagnosing brain cancer. The invention is useful in brain cancer therapy, treatment and diagnosis. The invention is also useful in gene therapy. The present sequence is human Gliten EST clone cDNA. This EST clone is used in the exemplification of the invention  
 Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;  
 Query Match 100.0%; score 22; DB 9; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Sq 1 AAGGTGGAGTCGAGGAGCTGC 22  
 5 AAGGTGGAGTCGAGGAGCTGC 26

RESULT 7  
 ID ABX95680  
 AC ABX95680 standard; DNA; 873 BP.  
 XX  
 AC ABX95680;  
 XX DT 30-JUN-2003 (first entry)  
 DE Human Gliten open reading frame.  
 XX Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;  
 KW Gliten; GMB.  
 XX Human gene encoding Gliten.  
 OS Homo sapiens.  
 PH Location/Qualifiers  
 FT 1..873  
 /!\*tgc= a  
 FT /partial  
 FT /product= "Gliten"  
 FT /note= "No stop codon shown"  
 XX US2003044811-A1.  
 PN 06-MAR-2003.  
 XX 20-OCT-2001; 2001US-00051769.  
 XX 20-OCT-2000; 2000US-0242160P.  
 XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 PA  
 PI Mckinnon RD;  
 DR WPI; 2003-418995/39.  
 PT New isolated nucleic acid representing a gene product associated with PT Glioblastoma Multiforme, designated as Gliten, useful as a molecular PT marker for tumor cell identification and classification.  
 PT XX Disclosure; Page 7; 11pp; English.  
 XX The invention relates to an isolated nucleic acid representing a gene CC

CC product associated with Glioblastoma Multiforme, designated as Gliten, CC comprising the human EST (expressed sequence tag) appearing as ABX95678, CC or a sequence that hybridises under stringent conditions to the EST, or CC its complement. Also included are a probe for use in identifying a CC patient at risk for progression into the malignant phenotype, comprising CC the nucleic acids detailed above, detecting whether a patient is at risk CC for progression into Glioblastoma Multiforme (comprising: (a) providing a CC sample from a patient; (b) adding the probe to the sample or performing CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) CC analysing levels of mRNA bound with the probe; and (d) treating a control CC sample to assess the level of mRNA in a control sample, where the CC presence of increased levels of mRNA expression in the sample in an CC amount higher than the control sample indicates risk for progression into CC Glioblastoma Multiforme), and a kit for use in detecting whether a CC patient is at risk for progression into Glioblastoma Multiforme (comprising the probe or primers). The probe or the kit is useful for CC detecting whether a patient is at risk for progression into Glioblastoma CC Multiforme (a type of brain cancer). The EST useful as molecular marker CC for tumour cell identification and classification and for diagnosing or CC identifying candidates at risk for progression into a malignant phenotype CC especially in brain cancer therapy, treatment and diagnosis. The present CC sequence is the open reading frame from the Gliten gene  
 XX SQ Sequence 873 BP; 137 A; 304 C; 279 G; 153 T; 0 U; 0 Other;  
 XX SQ Query Match 100.0%; Score 22; DB 7; Length 873;  
 Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAGGTGGAGTCGAGGAGCTGC 22  
 Db 370 AAGGTGGAGTCGAGGAGCTGC 391

RESULT 8  
 ID ABX95678  
 AC ABX95678;  
 XX  
 XX 30-JUN-2003 (first entry)  
 DE Human gene encoding Gliten.  
 XX Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;  
 KW Gliten; GMB.  
 XX Human gene encoding Gliten.  
 OS Homo sapiens.  
 PH Location/Qualifiers  
 FT 30..906  
 /!\*tgc= a  
 /product= "Gliten"  
 /note= "No stop codon shown"  
 XX US2003044811-A1.  
 PN 06-MAR-2003.  
 XX 20-OCT-2001; 2001US-00051769.  
 XX 20-OCT-2000; 2000US-0242160P.  
 XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 PA  
 PI Mckinnon RD;  
 DR WPI; 2003-418995/39.  
 PT New isolated nucleic acid representing a gene product associated with PT Glioblastoma Multiforme, designated as Gliten, useful as a molecular PT marker for tumor cell identification and classification.  
 PT XX Disclosure; Page 7; 11pp; English.  
 PS Example 2; Fig 1; 11pp; English.

The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLITEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into Glioblastoma Multiforme), and a kit for use in detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into Glioblastoma Multiforme (a type of brain cancer). The EST useful as molecular marker for tumour cell identification and classification and for diagnosing or identifying candidates at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is a region of human chromosome 10 (q25) containing the GLITEN

```

Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;
Query Match 100.0%; Score 22; DB 7; Length 960;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
          1 AAGTGGAGTTCGAGGAGCTGC 22
          ||||| | | | | | | | | | | |
          400 AAGTGGAGTTCGAGGAGCTGC 421

RESULT 9
D61559 standard; cdNA; 960 BP.
AAD61559;

AAD61559;
15-JAN-2004 (first entry)
Human GLTEN partial cDNA.

Human; glioblastoma multiforme; GBM; GLTEN; brain cancer; diagnosis;
gene therapy; chromosome 10; gene: ss 55.
```

Homo sapiens.  
US2003108915-A1.  
12-JUN-2003.  
20-AUG-2002; 2002US-00224624.  
20-OCT-2000; 2000US-0242160P.  
20-OCT-2001; 2001US-00051769.  
(UPNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
McKinnon RD;  
WPI; 2003-810848/76.  
Novel glioblastoma multiforme associated protein GLITN useful for  
treating glioblastoma multiforme and diagnosing brain cancer.  
Example 2; Fig 1; OPP; English.  
The present invention provides novel glioblastoma multiforme (GBM)

**RESULT 10**  
 ADE07789  
 ID ADE07789 standard; DNA; 1092 BP.  
 XX ADE07789;  
 XX AC  
 DT 29-JAN-2004 (first entry)  
 XX Novel coding sequence (useful for identifying genetic disorders) #855.  
 XX novel gene; novel protein; tissue marker; molecular weight marker;  
 XX chromosome marker; genetic disorder; gene; ds.  
 XX Unidentified.  
 OS  
 XX WO2003054152-A2.  
 XX PD  
 XX 03-JUL-2003.  
 XX PF  
 10-DEC-2002; 2002WO-US039555.  
 XX PR 10-DEC-2001; 2001US-0339739P.  
 PR 11-DEC-2001; 2001US-0339433P.  
 PR 14-MAR-2002; 2002US-0365091P.  
 PR 14-MAR-2002; 2002US-0365384P.  
 PR 12-APR-2002; 2002US-0372381P.  
 PR 12-APR-2002; 2002US-0372615P.  
 PR 22-APR-2002; 2002US-00128558.  
 PR 24-APR-2002; 2002US-0376045P.  
 XX PA  
 XX (HYSE-1) HYSEO INC.  
 PA

AA Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
XX  
WPI; 2003-569235/53.  
DR P-PSDE; ADE08700.  
XX  
PT New polynucleotides, useful for expressing recombinant proteins for  
PT analysis, characterization or therapeutic use, or as markers for tissues  
PT in which the corresponding protein is preferentially expressed.  
XX  
PS SEQ ID NO 855; 117pp; English.  
XX  
CC The invention comprises the amino acid and coding sequences of novel  
CC proteins. The DNA and protein sequences of the invention are useful as:  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present DNA sequence represents a gene of the  
CC invention.  
XX  
SQ Sequence 1092 BP; 190 A; 377 C; 322 G; 203 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 9; Length 1092;  
 Best Local Similarity 100.0%; Pred. No. 6.2.;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

<p>Qy 1 AAGGGGAGTTCGAGGAGCTGC 22</p> <p>Db 370 AAGGTGGAGTTGAGGAGCTGC 391</p>	<p>RESULT 11        ID ADB62530 standard; cDNA; 2639 BP.        ID ADB62530;        DT 04-DEC-2003 (first entry)        DE Human cDNA encoding clone HHDP20088160.        KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;        KW tissue regeneration; cell regeneration; membrane protein;        KW signal transduction-related protein; transcription-related protein;        KW osteoporosis; neurological disease; cancer; tumour.        OS Homo sapiens.        FH Key Location/Qualifiers        FT CDS 214 .2639        FT /tag= a        FT /partial        FT /product= "Clone HHDP20088160 protein"        XX EP1308459-A2.        XX PD 07-MAY-2003.        XX PR 28-MAR-2002; 2002EP-00007401.        XX PR 05-NOV-2001; 2001JP-00379298.        XX PR 25-JAN-2002; 2002US-00350978.        XX PA (HELI-) HELIX RES INST.        PA (REAS-) RES ASSOC BIOTECHNOLOGY.        PI Isogai T, Sugiyama T, Otsuki T, Wakanatsu A, Sato H, Ishii S;        PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;        PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;        DR WPI: 2003-450961/43.        DR P-PSDB; ADB64500.        XX PT New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.        XX Claim 1; Page; 222pp; English.</p>
--	---

The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesising the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them for developing a diagnostic marker or medicines for regulation of the expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell

regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumors. The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a cDNA of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

XX Sequence 2639 BP; 657 A; 702 C; 719 G; 561 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 9; Length 2639;  
 Best Local Similarity 100.0%; Pred. No. 6.5;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

<p>Qy 1 AAGGTGGAGTTCCAGGAGCTGC 22</p> <p>Db 583 AAGGTGGAGTTCCAGGAGCTGC 604</p>	<p>RESULT 12        ID AAD61564 standard; cDNA; 3832 BP.        AC AAD61564;        XX DT 15-JAN-2004 (first entry)        XX DS Human GLITEN cDNA.        XX Human; glioblastoma multiforme; GBM; GLITEN; brain cancer; diagnosis;        KW gene therapy; chromosome 10; gene; ss.        XX OS Homo sapiens.        XX FH Key Location/Qualifiers        FT CDS 178 .3642        FT /*tag= a        FT /product= "Human GLITEN protein"        FT /note= "The CDS is specifically claimed in claim 1"        FT primer_bind 1011. .1032        FT /*tag= b        FT /bound_moiety= "5' primer #843"        FT primer_bind 1405. .1429        FT /*tag= c        FT /bound_moiety= "5' primer #1405"        FT primer_bind 2083. .2108        FT /*tag= d        FT /bound_moiety= "5' primer #2083"        FT primer_bind 2356. .2378        FT /*tag= e        FT /bound_moiety= "5' primer #2355"        FT primer_bind 3625. .3650        FT /*tag= f        FT /bound_moiety= "3' primer #3616"        XX US2003108915-A1.        XX PD 12-JUN-2003.        XX PR 20-AUG-2002; 2002US-00224624.        XX PR 20-OCT-2000; 2000US-0242160P.        PR 20-OCT-2001; 2001US-00051169.        XX PA (UINE-) UNIV NEW JERSEY MEDICINE &amp; DENTISTRY.        XX McKinnon RD;        PI XX        DR XX        WPI: 2003-810848/76.        P-PSDB; ABK00758.</p>
--	---

Novel glioblastoma multiforme associated protein GLITEN useful for

PT treating glioblastoma multiforme and diagnosing brain cancer.  
XX Claim 1; Fig 4; Opp; English.

CC The present invention provides novel glioblastoma multiforme (GBM) associated protein GLI1EN useful for treating glioblastoma multiforme and diagnosing brain cancer. The invention is useful in brain cancer therapy, treatment and diagnosis. The invention is also useful in gene therapy. The present sequence is human GLI1EN cDNA. The human GLI1EN gene is located on chromosome 10.

XX Sequence 3832 BP; 1090 A; 926 C; 957 G; 859 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 9; Length 3832;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGTGGAGTTCGAGGAGCTGC 22  
Db 547 AAGTGGAGTTCGAGGAGCTGC 568

RESULT 13  
AAK51608  
ID AAK51608 standard; cDNA; 4450 BP.  
XX  
AC AAK51608;  
DT 06-NOV-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 153.  
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
XX Homo sapiens.  
OS Homo sapiens.  
XX WO200157190-A2.  
XX 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US004098.  
XX PR 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
PR 20-JUN-2000; 2000US-00598075.  
PR 19-JUL-2000; 2000US-00620325.  
PR 01-SEP-2000; 2000US-00654336.  
PR 15-SEP-2000; 2000US-00663561.  
PR 20-OCT-2000; 2000US-00693325.  
PR 30-NOV-2000; 2000US-00728422.  
XX (HYSEQ -) HYSEQ INC.

PA Tang YT, Liu C, Dumanac RT, Asuandi V, Zhou P, Xu C, Cao Y, Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW, Xue AU, Yang Y, Wejhrman T, Goodrich R,  
XX DR WPI; 2001-476283/51.  
DR P-PSDB; AAM78475.

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy.  
XX Claim 1; Page 862-867; 6221pp; English.

PS The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM7832-AAM80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The nucleotides and polynucleotides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities, e.g.; stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and/or activin/inhibitin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3566 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

CC XX Sequence 4450 BP; 1252 A; 987 C; 1047 G; 1164 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 22; DB 4; Length 4450;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGTGGAGTTCGAGGAGCTGC 22  
Db 370 AAGTGGAGTTCGAGGAGCTGC 391

RESULT 14  
ABA09174  
ID ABA09174 standard; cDNA; 4470 BP.  
XX  
AC ABA09174;  
DT 11-JAN-2002 (first entry)  
XX Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.

DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor; haematopoiesis regulation; tissue growth; immunomodulator; activin; KW inhibitor; chemotaxis; chemokinesis; thrombolytic; oncogenesis; proliferation; metastasis; cancer; tumour; hematopoietic disorder; myeloid cell disorder; lymphoid cell disorder; asthma; arthritis; chronic inflammatory condition; proinflammatory; retinopathy; atherosclerosis; coronary heart disease; arterial ischaemia; bone disorder; osteoporosis; vascular growth disorder; tissue regeneration; wound healing; infection; immune disorder; cell culture; drug screening; gene therapy; antiinflammatory; antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic; cytostatic; osteopathic; vasotropic; cardiotonic; virucide; antibacterial; antifungal; pulmonary; antiluler; ss.

XX (HYSEQ -) HYSEQ INC.

PA Homo sapiens.  
XX PN WO200157188-A2.  
AX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US003800.  
XX PR 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
PR 20-JUN-2000; 2000US-00598075.  
PR 19-JUL-2000; 2000US-00620325.  
PR 01-SEP-2000; 2000US-00654336.  
PR 15-SEP-2000; 2000US-00663561.  
PR 20-OCT-2000; 2000US-00693325.  
PR 30-NOV-2000; 2000US-00728422.  
XX DR WPI; 2001-457740/49.  
PR 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.

XX (HYSEQ -) HYSEQ INC.

PA Tang YT, Liu C, Dumanac RT;  
PI Tang YT, Liu C, Dumanac RT;  
XX DR WPI; 2001-457740/49.  
PR 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.

XX (HYSEQ -) HYSEQ INC.

PA Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer.

XX Claim 1; Page 813-814; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and sequences ABA08225-RBA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides

or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haemopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activating or inhibiting related activities; chemokine or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention.

XX Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 4; Length 4470;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACGTGGAGTTGAGGAAGCTGC 22  
Db 370 AACGTGGAGTTGAGGAAGCTGC 391

RESULT 15

AAK52592  
ID AAK52592 standard; cDNA; 4470 BP.  
XX  
AC AAK52592;  
DT 06-NOV-2001 (first entry)

XX DE Human Polynucleotide SEQ ID NO 2121.  
XX  
PR 05-FEB-2001; 2001IWO-US004098.

XX PR 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
PR 20-JUN-2000; 2000US-00598075.  
PR 19-JUL-2000; 2000US-00620325.  
PR 01-SEP-2000; 2000US-00654936.  
PR 15-SEP-2000; 2000US-00663561.

CC PR 20-OCT-2000; 2000US-00693325.  
CC PR 30-NOV-2000; 2000US-00728422.

XX PA (HYSE-) HYSEQ INC.

CC XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y; PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW; PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX DR WPI; 2001-476283/51.  
XX DR P-PSDB; AAM79459.  
XX PT Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy.  
XX PS Claim 1; Page 4496-4497; 6221pp; English.  
XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM8023-AAM802) that exhibit activity relating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haemopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activation/inhibition activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM8020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

XX SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 4; Length 4470;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGTGGAGTTGAGGAAGCTGC 22  
Db 370 AGGTGGAGTTGAGGAAGCTGC 391

Search completed: July 18, 2004, 12:45:04  
Job time : 38.7148 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:32:20 ; Search time 7.42951 Seconds  
(without alignments)  
1643.301 Million cell updates/sec

Title: US-10-051-769-5  
Perfect score: 22  
Sequence: 1 aaggtagatccgaggagtcg 22

Scoring table: IDENTITY\_NUC  
Gapop 10<sup>-0</sup>, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:  
 1: /cgn2\_6\_ptodata/2/ina/5A\_COMB.seq;\*  
 2: /cgn2\_6\_ptodata/2/ina/5B\_COMB.seq;\*  
 3: /cgn2\_6\_ptodata/2/ina/5C\_COMB.seq;\*  
 4: /cgn2\_6\_ptodata/2/ina/5D\_COMB.seq;\*  
 5: /cgn2\_6\_ptodata/2/ina/5CTUS\_COMB.seq;\*  
 6: /cgn2\_6\_ptodata/2/ina/backfile1.seq;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	4303	4	US-09-976-994-899 Sequence 899, Appl
2	17.8	80.9	2184	3	US-09-307-265A-2 Sequence 2, Appl
3	17.2	78.2	1332	2	US-08-481-814A-3 Sequence 3, Appl
4	17.2	78.2	1489	3	US-08-836-582-1 Sequence 1, Appl
5	17.2	78.2	1489	4	US-09-265-566-1 Sequence 1, Appl
6	17.2	78.2	1489	4	US-09-242-137-3 Sequence 3, Appl
7	17.2	78.2	1489	4	US-09-023-655-871 Sequence 871, Appl
8	17.2	78.2	2886	4	US-09-221-017B-963 Sequence 863, Appl
9	17.2	78.2	3924	3	US-08-726-214-9 Sequence 9, Appl
10	16.8	76.4	336	4	US-09-821-167-5 Sequence 5, Appl
11	16.8	76.4	735	3	US-08-998-416-880 Sequence 880, Appl
12	16.8	76.4	1322	4	US-08-973-005A-1 Sequence 1, Appl
13	16.8	76.4	1347	4	US-09-016-1261 Sequence 1261, Appl
14	16.8	76.4	1650	2	US-08-436-771-10 Sequence 10, Appl
15	16.8	76.4	1650	2	US-08-434-998-10 Sequence 10, Appl
16	16.8	76.4	1650	2	US-08-487-797-10 Sequence 10, Appl
17	16.8	76.4	1650	5	FCT-US95-02058-10 Sequence 103, Appl
18	16.8	76.4	2213	4	US-09-620-212D-1038 Sequence 1, Appl
19	16.8	76.4	2214	3	US-08-864-038A-1 Sequence 10, Appl
c	20	16.8	76.4	2374	3 US-09-347-001-3 Sequence 3, Appl
21	16.8	76.4	3331	3	US-08-864-038A-2 Sequence 2, Appl
22	16.8	76.4	3331	3	US-08-864-038A-4 Sequence 4, Appl
23	16.8	76.4	5325	4	US-10-164-595-17 Sequence 17, Appl
24	16.8	76.4	5334	4	US-10-164-595-21 Sequence 21, Appl
25	16.8	76.4	5581	4	US-10-164-595-19 Sequence 19, Appl
26	16.8	76.4	11188	4	US-09-821-167-1 Sequence 1, Appl
27	16.4	74.5	2020	4	US-09-620-312D-751 Sequence 751, Appl

**ALIGNMENTS**

RESULT 1  
US-09-976-594-899  
; Sequence 899, Application US/09976594  
; Patent No. 6673449  
; GENERAL INFORMATION:  
; APPLICANT: Buchbinder, Jenny  
; INVENTOR: Furness, Michael  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041.US  
; CURRENT APPLICATION NUMBER: US/09/976, 594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240, 409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO: 899  
; LENGTH: 4303  
; FEATURE:  
; NAME KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1  
US-09-976-594-899  
Query Match 100.0%; Score 22; DB 4;  
Best Local Similarity 100.0%; Pred. No. 0.83;  
Matches 22; Conservative 0; Mismatches 0;  
Indels 0; Gaps 0;

QY 1 AAGGTGAGTTGGAGGCTGC 22  
Db 83 AAGGTGAGTTGGAGGCTGC 104

RESULT 2  
US-09-307-265A-2  
; Sequence 2, Application US/09307265A  
; Patent No. 6225456  
; GENERAL INFORMATION:  
; APPLICANT: Gu, Trent  
; INVENTOR: Orita, Satoshi  
; FILE REFERENCE: UTC-03732  
; CURRENT APPLICATION NUMBER: US/09/307, 265A  
; CURRENT FILING DATE: 1999-05-06  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 2  
; LENGTH: 2184  
; TYPE: DNA

ORGANISM: Homo sapiens  
US-09-307-265A-2

Query Match 3 Score 17.8; DB 3; Length 2184;  
Best Local Similarity 90.5%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGGTGGAGTTGAGGAGCTGC 22  
Db 797 AGCTGGAGTTCGAGCAGTCG 817

RESULT 3  
US-08-481-814A-3  
; Sequence 3, Application US/08481814A  
; Patent No. 5865040

GENERAL INFORMATION:  
APPLICANT: Hsu, Yen-Ming TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Biogen, Inc.  
STREET: 14 Cambridge Center  
CITY: Cambridge  
STATE: Massachusetts  
ZIP: 02142

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,582  
FILING DATE: 14-MAY-1997  
CLASSIFICATION: 425  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/00866  
FILING DATE: 18-APR-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: GB 9423049.7  
FILING DATE: 15-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Arthur R. Crawford  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 620-18  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
NAME/KEY: CDS  
LENGTH: 1489 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 27..1268  
US-08-836-582-1

Query Match 1 Score 78.2%; Best Local Similarity 86.4%; Pred. No. 93;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AAGGTGGAGTTGAGGAGCTGC 22  
Db 327 AAGGCAGAGATCGAGGAGCTGC 348

RESULT 5  
US-09-265-556-1  
; Sequence 1, Application US/09265556  
; Patent No. 6303335

GENERAL INFORMATION:  
APPLICANT: Bernards, Rene  
TITLE OF INVENTION: Transcription factor E2F-4  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vanderhye  
STREET: 1100 No. 630335th Glebe Road, 8th Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: US  
ZIP: 22201-4714

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/265,566  
 FILING DATE: 10-Mar-1999  
 CLASSIFICATION: <Unknown>  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/836,582  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arthur R. Gifford  
 REGISTRATION NUMBER: 25,327  
 REFERENCE/DOCKET NUMBER: 620-18  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1489 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 27..1268  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-265-566-1

Query Match 78.2%; Score 17.2; DB 4; Length 1489;  
 Best Local Similarity 86.4%; Pred. No. 93;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACGGAGGTCTAGGAGCTGC 22  
 Db 327 AAGGCAAGATCGAGGAGCTGC 348

RESULT 6  
 US-09-242-737-3  
 Sequence 3 Application US/09242737  
 Patent No. 6368809

GENERAL INFORMATION:  
 APPLICANT: BERNARDS, REN (UNNAMED INVENTOR)  
 TITLE OF INVENTION: E2F UBIQUITINATION DOMAIN, AND ASSAYS FOR  
 INHIBITORS OF E2F UBIQUITINATION  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: NIXON & VANDERHYE P.C.  
 STREET: 1100 NORTH GLEBE ROAD  
 CITY: ARLINGTON  
 STATE: VIRGINIA  
 COUNTRY: U.S.A.  
 ZIP: 22201-4714

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/242,737  
 FILING DATE: 23-Feb-1999  
 CLASSIFICATION: <Unknown>  
 APPLICATION NUMBER: PCT/GB97/02293  
 FILING DATE: 22-AUG-1997  
 APPLICANT NUMBER: GB 9617697.9  
 FILING DATE: 23-AUG-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: WILSON, MARY J.  
 REGISTRATION NUMBER: 32,955  
 REFERENCE/DOCKET NUMBER: 620-63  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 816-4000  
 TELEX/FAX: (703) 816-4100  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:

LENGTH: 1489 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 27..1268  
 SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
 US-09-242-737-3

Query Match 78.2%; Score 17.2; DB 4; Length 1489;  
 Best Local Similarity 86.4%; Pred. No. 93;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACGGAGGTCTAGGAGCTGC 22  
 Db 327 AAGGCAAGATCGAGGAGCTGC 348

RESULT 7  
 US-09-023-655-871  
 Sequence 871 Application US/09023655  
 Patent No. 6607879

GENERAL INFORMATION:  
 APPLICANT: Cockes, Benjamin G.  
 APPLICANT: Susan G. Stuart  
 APPLICANT: Jeffrey J. Sellhamer  
 TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
 TITLE OF INVENTION: EXPRESSION  
 NUMBER OF SEQUENCES: 1508  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 STREET: 3174 PORTER DRIVE  
 CITY: PALO ALTO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94304

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/023,655  
 FILING DATE: HEREWITH  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Zeller, Karen J.  
 REGISTRATION NUMBER: 37,071  
 REFERENCE/DOCKET NUMBER: PA-0001 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 855-0555  
 TELEFAX: (650) 845-4166  
 INFORMATION FOR SEQ ID NO: 871:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1489 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: GENBANK  
 CLOONE: GI061145

US-09-023-655-871

Query Match 78.2%; Score 17.2; DB 4; Length 1489;  
 Best Local Similarity 86.4%; Pred. No. 93;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACGGAGGTCTAGGAGCTGC 22

RESULT 9  
US-08-726-214-9  
Sequence 9, Application US/08/726214  
; Patent No. 6107076

GENERAL INFORMATION:  
 APPLICANT: Gilman, Alfred G.  
 TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE  
 NUMBER OF SEQUENCES: 31  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Arnold, White & Durkee  
 STREET: P.O. Box 4433  
 CITY: Houston  
 STATE: Texas  
 COUNTRY: United States of America  
 ZIP: 77210

COMPUTER READABLE FORM:  
 COMPUTER: IBM PC compatible  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/726,214  
 FILING DATE: Concurrently Herewith  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/005,498  
 FILING DATE: 04-OCT-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Highlander, Steven L.  
 REGISTRATION NUMBER: 37,642  
 REFERENCE/DOCKET NUMBER: US/450  
 TELECOMMUNICATION INFORMATION:  
 PHONE: (512) 418-3000  
 TELEFAX: (512) 474-7577  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3,924 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-726-214-9

Query Match 78.2%; Score 17.2; DB 3; Length 3924;  
 Best Local Similarity 86.4%; Pred. No. 1e+02; 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 10  
US-09-821-167-5  
Sequence 5, Application US/09821167  
; Patent No. 6569668

GENERAL INFORMATION:  
 APPLICANT: Hosted Jr., Thomas J.  
 TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria  
 TITLE OF INVENTION: Plasmid pMR2 and Vectors Made Therefrom  
 FILE REFERENCE: IN01149Q

CURRENT APPLICATION NUMBER: US/09/821,167  
 CURRENT FILING DATE: 2004-03-29  
 PRIOR APPLICATION NUMBER: US 60/194,461  
 PRIOR FILING DATE: 2000-04-04  
 NUMBER OF SEQ ID NOS: 15  
 SOFTWARE: Patent In Ver. .2.1  
 SEQ ID NO 5  
 LENGTH: 336  
 TYPE: DNA  
 ORGANISM: Micromonospora rosaria

Query Match 78.2%; Score 17.2; DB 4; Length 2886;  
 Best Local Similarity 86.4%; Pred. No. 99; 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 11  
US-09-221-017B-863  
Sequence 11, Application US/09/221017B  
; Patent No. 6447499

GENERAL INFORMATION:  
 APPLICANT: Ross, Bruce C.  
 TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
 NUMBER OF SEQUENCES: 1120  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: MORRISON & FOERSTER  
 STREET: 755 PAGE MILL ROAD  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 9404-1018

COMPUTER READABLE FORM:  
 COMPUTER: IBM Compatible  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows  
 SOFTWARE: FastSEQ for Windows Version 2.0b  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/221,017B  
 FILING DATE: 23-DEC-1998  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP1182  
 FILING DATE: 31-DEC-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP1546  
 FILING DATE: 30-JAN-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PP2911  
 FILING DATE: 09-APR-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/AU98/01023  
 FILING DATE: 10-DEC-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Monroy, Gladys H  
 REGISTRATION NUMBER: 32,430  
 REFERENCE/DOCKET NUMBER: 27340-20021.00  
 TELECOMMUNICATION INFORMATION:  
 PHONE: 650-813-5600  
 TELEFAX: 650-494-0792  
 TELEX: 706141  
 LOCATION: 1...2886  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2886 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: UNKNOWN  
 ORIGINAL SOURCE:  
 ORGANISM: PORPHYROMONAS GINGIVALIS  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 1...2886  
 US-09-221-017B-863

Query Match 78.2%; Score 17.2; DB 4; Length 2886;  
 Best Local Similarity 86.4%; Pred. No. 99; 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 12  
US-08-726-214-9  
Sequence 12, Application US/08/726214  
; Patent No. 6107076

GENERAL INFORMATION:  
 APPLICANT: Horan, Ann C.  
 TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria  
 TITLE OF INVENTION: Plasmid pMR2 and Vectors Made Therefrom  
 FILE REFERENCE: IN01149Q

CURRENT APPLICATION NUMBER: US/09/821,167  
 CURRENT FILING DATE: 2004-03-29  
 PRIOR APPLICATION NUMBER: US 60/194,461  
 PRIOR FILING DATE: 2000-04-04  
 NUMBER OF SEQ ID NOS: 15  
 SOFTWARE: Patent In Ver. .2.1  
 SEQ ID NO 5  
 LENGTH: 336  
 TYPE: DNA  
 ORGANISM: Micromonospora rosaria

Query Match 78.2%; Score 17.2; DB 22; Length 887;  
 Best Local Similarity 86.4%; Pred. No. 99; 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 13  
US-08-726-214-9  
Sequence 13, Application US/08/726214  
; Patent No. 6107076

GENERAL INFORMATION:  
 APPLICANT: Horan, Ann C.  
 TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria  
 TITLE OF INVENTION: Plasmid pMR2 and Vectors Made Therefrom  
 FILE REFERENCE: IN01149Q

CURRENT APPLICATION NUMBER: US/09/821,167  
 CURRENT FILING DATE: 2004-03-29  
 PRIOR APPLICATION NUMBER: US 60/194,461  
 PRIOR FILING DATE: 2000-04-04  
 NUMBER OF SEQ ID NOS: 15  
 SOFTWARE: Patent In Ver. .2.1  
 SEQ ID NO 5  
 LENGTH: 336  
 TYPE: DNA  
 ORGANISM: Micromonospora rosaria

Query Match 78.2%; Score 17.2; DB 22; Length 887;  
 Best Local Similarity 86.4%; Pred. No. 99; 0; Mismatches 3; Indels 0; Gaps 0;

US-09-821-167-5  
 Query Match 76.4%; Score 16.8; DB 4; Length 336;  
 Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGGTGGAGTTGAGGAGCTG 21  
 Db 209 AGGTGGAGTTGAGGAGCTG 228

RESULT 11  
 US-08-998-416-880  
 Sequence 880, Application US/08998416  
 ; Patent No. 6239264  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Philippsen, Peter  
 ; APPLICANT: Pohlmann, Rainer  
 ; APPLICANT: Steinher, Sabine  
 ; APPLICANT: Mohr, Christine  
 ; APPLICANT: Wendland, Jurgen  
 ; APPLICANT: Knechtle, Philipp  
 ; APPLICANT: Rebischung, Corinne  
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHYA GOSSYPII  
 ; TITLE OF INVENTION: AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 1152  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 6239264artis Corporation  
 ; STREET: 3054 Cornwallis Road  
 ; CITY: Research Triangle Park  
 ; STATE: No. 6239264th Carolina  
 ; COUNTRY: USA  
 ; ZIP: 27709  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/973,005A  
 ; FILING DATE: 01-Dec-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 133741/1995  
 ; FILING DATE: 31-MAY-1995  
 ; APPLICATION NUMBER: JP 285401/1995  
 ; FILING DATE: 01-NOV-1995  
 ; APPLICATION NUMBER: JP 52010/1996  
 ; FILING DATE: 08-MAR-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Bent, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 081356/0113  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1322 base pairs  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Human  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-08-973-005A-1

US-08-973-005A-1  
 Query Match 76.4%; Score 16.8; DB 4; Length 1322;  
 Best Local Similarity 90.0%; Pred. No. 1.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GGTCGAGCTTCGAGGAGCTG 22  
 Db 465 GGTCGAGCTTCGATGTGCTGC 484

RESULT 13  
 US-09-016-444-1261  
 Query Match 76.4%; Score 16.8; DB 3; Length 735;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AGGTGGAGTTGAGGAGCTG 21  
 Db 114 AGGTGGAGTTGAGGAGCTG 133

RESULT 12  
 US-10-051-769-5.std.rni

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 STREET: 3174 PORTER DRIVE  
 CITY: PALO ALTO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94304

COMPUTER READABLE FORM:  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/016,434  
 FILING DATE: HEREWTH

PRIOR APPLICATION DATA:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Zeller, Karen J.  
 REGISTRATION NUMBER: 37,071  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 845-4166  
 TELEFAX: (650) 845-0555

INFORMATION FOR SEQ ID NO: 1261:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1347 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

IMMEDIATE SOURCE:  
 LIBRARY: GENBANK  
 CLONE: 9220080

US-09-016-434-1261

Query Match Score 16.8; DB 4; Length 1347;  
 Best Local Similarity 90.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;

Qy 3 GGTGGAGTCGAGGAGCTGC 22  
 Db 495 GGTGGAGTCGATGCTGC 514

RESULT 14  
 US-08-436-771-10  
 Sequence 10, Application US/08436771  
 GENERAL INFORMATION:  
 Patent No. 5861300

ATTORNEY/AGENT INFORMATION:  
 APPLICANT: Silverman, Robert H.  
 ADDRESSEE: SenGupta, Debjendu N.  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0., Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/434,998  
 FILING DATE:  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 305/527/2498  
 FILING DATE: 18-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Manso, Peter J.  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0., Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/436,771  
 FILING DATE:  
 CLASSIFICATION: 536

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/198,973  
 FILING DATE: 18-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Manso, Peter J.  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 305/527/2498  
 FILING DATE: 18-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Manso, Peter J.  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0., Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/434,998  
 FILING DATE:  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 305/527/2498  
 FILING DATE: 18-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Manso, Peter J.  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0., Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/436,771  
 FILING DATE:  
 CLASSIFICATION: 536

RESULT 15  
 US-08-434-998-10  
 Sequence 10, Application US/08434998  
 GENERAL INFORMATION:  
 Patent No. 5866781

GENERAL INFORMATION:  
 APPLICANT: Silverman, Robert H.  
 ADDRESSEE: SenGupta, Debjendu N.  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0., Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/434,998  
 FILING DATE:  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 305/527/2498  
 FILING DATE: 18-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Manso, Peter J.  
 STREET: 200 E. Broward Boulevard  
 CITY: Fort Lauderdale  
 STATE: Florida  
 COUNTRY: USA  
 ZIP: 33301

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0., Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/436,771  
 FILING DATE:  
 CLASSIFICATION: 536

Query Match 76.4%; Score 16.8; DB 2;  
Best Local Similarity 90.0%; Pkd. No. 1.e+02;  
Matches 18; Conservative 0; Mismatches 2;  
Indels 0; Gaps 0;  
Qy 3 GGTGGAGTCGAGAGCTGC 22  
Db 526 GGCGAGTCGATGTGCTGC 545

Search completed: July 18, 2004, 14:38:29  
Job time : 8.42951 secs



Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	15	US-10-051-769-5 Sequence 5, Appli
2	22	100.0	22	15	US-10-051-769-5 Sequence 5, Appli
3	22	100.0	261	15	US-10-051-769-2 Sequence 2, Appli
4	22	100.0	261	15	US-10-051-769-2 Sequence 2, Appli
5	22	100.0	263	15	US-10-051-769-4 Sequence 4, Appli
6	22	100.0	263	15	US-10-051-769-4 Sequence 4, Appli
7	22	100.0	873	15	US-10-051-769-3 Sequence 3, Appli
8	22	100.0	960	15	US-10-051-769-1 Sequence 1, Appli
9	22	100.0	960	15	US-10-051-769-1 Sequence 1, Appli
10	22	100.0	960	15	US-10-051-769-1 Sequence 68, Appli
11	22	100.0	2639	16	US-10-051-769-1 Sequence 9, Appli
12	22	100.0	3465	15	US-10-051-769-1 Sequence 63, Appli
13	22	100.0	3649	13	US-10-051-769-1 Sequence 7, Appli
14	22	100.0	3832	15	US-10-051-769-1 Sequence 2, Appli

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	15	US-10-051-769-5 Sequence 5, Appli
2	22	100.0	22	15	US-10-051-769-5 Sequence 5, Appli
3	22	100.0	261	15	US-10-051-769-2 Sequence 2, Appli
4	22	100.0	261	15	US-10-051-769-2 Sequence 2, Appli
5	22	100.0	263	15	US-10-051-769-4 Sequence 4, Appli
6	22	100.0	263	15	US-10-051-769-4 Sequence 4, Appli
7	22	100.0	873	15	US-10-051-769-3 Sequence 3, Appli
8	22	100.0	960	15	US-10-051-769-1 Sequence 1, Appli
9	22	100.0	960	15	US-10-051-769-1 Sequence 1, Appli
10	22	100.0	2639	16	US-10-051-769-1 Sequence 68, Appli
11	22	100.0	3465	15	US-10-051-769-1 Sequence 9, Appli
12	22	100.0	3649	13	US-10-051-769-1 Sequence 63, Appli
13	22	100.0	3832	15	US-10-051-769-1 Sequence 7, Appli

RESULT 2  
US-10-051-769-5  
; Sequence 5, Application US/100517699  
; Publication No. US2003044811A1  
; GENERAL INFORMATION:  
; APPLICANT: MCKINNON, Randy D.  
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
; CURRENT APPLICATION NUMBER: US/10-051-769-5  
; FILE REFERENCE: 268-260 (RWJ-00-37)  
; CURRENT FILING DATE: 2001-10-20  
; PRIOR APPLICATION NUMBER: US 60/242,160  
; PRIOR FILING DATE: 2000-10-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 5  
; LENGTH: 22  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Primer  
; ;  
US-10-051-769-5  
; Sequence 1, Application US/10224624  
; Publication No. US20030108915A1  
; GENERAL INFORMATION:  
; APPLICANT: MCKINNON, Randall D.

**ALIGNMENTS**

```

; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLIEN
; FILE REFERENCE: 54704-8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIORITY NUMBER: 60/242,160
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 5
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE: OTHER INFORMATION: Synthetic Primer
; OTHER INFORMATION: Synthetic Primer
US-10-224-624-5
Query Match 100.0%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGTGGAGTTCCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCCGAGGAGCTGC 26
RESULT 5
US-10-051-769-4
Query Match 100.0%; Score 22; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGTGGAGTTCCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCCGAGGAGCTGC 26
RESULT 5
US-10-051-769-4
Query Match 100.0%; Score 22; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGTGGAGTTCCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCCGAGGAGCTGC 26
RESULT 5
US-10-051-769-4
Query Match 100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGTGGAGTTCCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCCGAGGAGCTGC 26
RESULT 6
US-10-224-624-4
Query Match 100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGTGGAGTTCCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCCGAGGAGCTGC 26
RESULT 6
US-10-224-624-4
Query Match 100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGTGGAGTTCCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCCGAGGAGCTGC 26
RESULT 6
US-10-224-624-4
Query Match 100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGTGGAGTTCCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCCGAGGAGCTGC 26
RESULT 6
US-10-224-624-4
Query Match 100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGTGGAGTTCCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCCGAGGAGCTGC 26

```

RESULT 7  
US-10-051-769-3  
Sequence 3, Application US/10051769  
Publication No. US20030044811A1  
GENERAL INFORMATION:  
APPLICANT: MCKINNON, Randy D.  
TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
FILE REFERENCE: 268/260 (RMJ-00-3)  
CURRENT APPLICATION NUMBER: US/10/051,769  
CURRENT FILING DATE: 2001-10-20  
PRIOR APPLICATION NUMBER: US 60/242,160  
PRIOR FILING DATE: 2000-10-20  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-10-051-769-3  
Query Match 100.0%; Score 22; DB 15; Length 873;  
Best Local Similarity 100.0%; Pred. No. 2.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AAGGTGGAGTTCGAGGAGCTGC 22  
Db 370 AAGGTGGAGTTCGAGGAGCTGC 391

RESULT 8  
US-10-224-624-3  
Sequence 3, Application US/10224624  
Publication No. US20030108915A1  
GENERAL INFORMATION:  
APPLICANT: MCKINNON, Randall D.  
TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN  
FILE REFERENCE: 54704.8059.US00  
CURRENT APPLICATION NUMBER: US/10/224,624  
CURRENT FILING DATE: 2002-08-20  
PRIOR APPLICATION NUMBER: 60/242,160  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 10/051,769  
PRIOR FILING DATE: 2001-10-20  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-10-224-624-3  
Query Match 100.0%; Score 22; DB 15; Length 960;  
Best Local Similarity 100.0%; Pred. No. 2.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AAGGTGGAGTTCGAGGAGCTGC 22  
Db 400 AAGGTGGAGTTCGAGGAGCTGC 421

RESULT 9  
US-10-051-769-1  
Sequence 1, Application US/10051769  
Publication No. US20030044811A1  
GENERAL INFORMATION:  
APPLICANT: MCKINNON, Randy D.  
TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
FILE REFERENCE: 268/260 (RMJ-00-3)  
CURRENT APPLICATION NUMBER: US/10/051,769  
CURRENT FILING DATE: 2001-10-20  
PRIOR APPLICATION NUMBER: US 60/242,160  
PRIOR FILING DATE: 2000-10-20



```
; LENGTH: 4470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-276-774-950

Query Match      100.0%; Score 22; DB 13; Length 4470;
Best Local Similarity 100.0%; Prod. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy   1 AAGGTGGACTCGAGGAGCTGC 22
Db   370 AAGTGGAGTCGAGGAGCTGC 391
```

Search completed: July 18, 2004, 14:47:31  
Job time : 38.5902 secs



GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:55:38 ; Search time 272.872 Seconds  
 (without alignments)  
 2407.604 Million cell updates/sec

Title: US-10-051-769-5  
 Perfect score: 22  
 Sequence: 1 aaggtagatccggaggatgc 22

Scoring table: IDENTITY\_NUC  
 Gapop 10\_0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:

55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : EST:\*

```
1: em_estba:*
2: em_estbum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_eston:*
16: em_eston:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_mam:*
22: em_gss_mus:*
23: em_gss_pro:*
24: em_gss_lod:*
25: em_gss_phg:*
26: em_gss_vrl:*
27: gb_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query	Match Length	DB ID	Description
1	22	100.0	437	13	BY276266
2	22	100.0	458	10	BB638573
c 3	22	100.0	583	10	BF330450
c 4	22	100.0	667	13	BY734796

5	22	100.0	702	13	BUT06646 UI-M-FOO-
6	22	100.0	723	9	AW013379
7	22	100.0	738	29	CC567140 CH240441
8	22	100.0	930	13	BQ960960 AGENCOURT
9	22	100.0	944	13	BX345030 BX345030
10	22	100.0	2052	11	AK041847 Mus muscu
11	22	100.0	3069	11	BCO33019 Homo sapi
12	20.4	92.7	681	10	BB866050 BB866050
13	20.4	92.7	730	13	BU49181 C037674032
14	20.4	92.7	898	13	BU453045 603219082
15	20	90.9	989	13	BX341552 BX341552
16	19.4	88.2	411	10	BE381105 WR0654 G
17	19	86.4	610	10	BE469647 Tphdk0206
18	19	86.4	910	10	BE469591 Tphdk0201
19	18.8	85.5	345	29	CC536807 CC536807
20	18.8	85.5	369	12	BI481867 RE6476 .5
21	18.8	85.5	404	12	BI361904 RE64311.5
22	18.8	85.5	412	9	AIS19791 LD39572.5
c	23	18.8	85.5	414	10 BF848453 PM4-EN006
c	24	18.8	85.5	425	13 BY260653 BY260653
c	25	18.8	85.5	426	10 BE84892 171707 BA
c	26	18.8	85.5	439	28 AZ048370 LMJFV1.1
c	27	18.8	85.5	445	9 AA820376 LD22957.5
c	28	18.8	85.5	448	9 AA246831 LD03821.5
c	29	18.8	85.5	456	9 AI518512 LD38052.5
c	30	18.8	85.5	462	12 BI365738 RE35111.5
c	31	18.8	85.5	473	12 BI228997 RE26679.5
c	32	18.8	85.5	503	12 BI172696 RE15344.5
c	33	18.8	85.5	504	12 BM256280 518617 MA
c	34	18.8	85.5	523	13 BU443584 604142115
c	35	18.8	85.5	524	10 BE236368 144133 MA
c	36	18.8	85.5	532	12 BI236723 RE12830.5
c	37	18.8	85.5	534	9 AI512754 LD147261.5
c	38	18.8	85.5	547	9 AV665832 AV665832
c	39	18.8	85.5	559	14 CP365905 836565 MA
c	40	18.8	85.5	573	12 BM487827 PGM2n.PK
c	41	18.8	85.5	576	12 BI539978 452874 MA
c	42	18.8	85.5	598	9 AI511790 LD43313.5
c	43	18.8	85.5	603	12 BI264150 RE49211.5
c	44	18.8	85.5	604	12 BI216164 RE23101.5
c	45	18.8	85.5	608	9 AIS163334 LD42168.5

### ALIGNMENTS

RESULT 1	BY276266	437 bp mRNA linear	EST 11-DEC-2002
LOCUS	BY276266 RIKEN full-length enriched, visual cortex	Mus musculus	
DEFINITION	CDNA clone K430318E15 5'	mRNA sequence.	
ACCESSION	BY276266		
VERSION	BY276266.1	GI: 26466603	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 437)		
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamamoto, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Shchrimi, L.M., Kanapin, A., Bult, C., Hume, D.A., Quackenbush, J., Hill, D.P., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Garibaldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanaai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagai, Y., A., Kurochkin, I.V., Lee, Y., Lenhoff, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, J., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,		

Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sepulve, C.A., Setou, M., Shimeda, K., Sultan, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, U., Whalesett, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Sakazume, N., Hirozane-Kishikawa, T., Waki, K., Kawaji, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imozumi, K., Ishii, Y., Itoh, M., Nagawa, T., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., and Hayashizaki, Y.

**TITLE** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
**JOURNAL** Nature 420, 563-573 (2002)  
**PUBLNED** 22354683  
**COMMENT** Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute of Physical and Chemical Research (RIKEN)  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp/  
[URL: http://genome.gsc.riken.go.jp/](http://genome.gsc.riken.go.jp/)  
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, T., Imotani, K., Ishii, Y., Itoh, M., Konno, H., Hirozane, T., Murata, M., Nakamura, K., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Muramatsu, M., and Hayashizaki, Y. Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384 format sequencing Pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Division of Experimental Animal Research in Riken contributed to preparation mouse tissues.  
 Tissues were provided by Michela Pagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute Genomic Sciences Center and Genome Science Research Group in Riken).  
 RIKEN 2-1 Hirosawa-Wako-ishi Saitama 351-0198 Japan ) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES**  
**source**

RESULT 2  
 BB638573 LOCUS BB638573 RIKEN full-length enriched, 3 days neonate mRNA linear EST 31-AUG-2001  
 DEFINITION BB638573 RIKEN full-length enriched, 3 days neonate thymus Mus  
 ACCESION BB638573  
 VERSION BB638573.1 GI:15401196  
 KEYWORDS EST  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 458)  
 AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hayashizaki, Y., Hiramoto, K., Hori, F., Itoh, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Saka, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takeda, Y., Tanaka, T., Toyai, T., Muramatsu, M., and Hayashizaki, Y.  
 TITLE RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9216  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp/  
 Carninci, P., Fukuda, S., Hayashizaki, Y., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 Sugahara, Y., Fujiwake, S., Inoue, K., Togawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuru, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384 format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Yamamoto, T., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y., and Hayashizaki, Y. Mapping of 15032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 Pre, L72-L86 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES**  
**source**

/organism="Mus musculus"  
/mol type="mRNA"  
/stran="C57BL/6J"  
/db\_xref="Taxon:10090"  
/clone="K430318S15"  
/tissue\_type="visual cortex"  
/clone\_id="RIKEN full-length enriched, visual cortex"

**ORIGIN**

Query Match 100.0% Score 22; DB 13; Length 437;  
 Best Local Similarity 100.0% Pred. No. 3.9e+02;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGTGGAGTTCAGGACTGC 22  
 Db 85 AAGGTGGAGTTCAGGAGCTGC 106

**Location/Qualifiers**

1..437  
 /mol type="mRNA"  
 /organism="Mus musculus"  
/mol type="mRNA"  
/db\_xref="Taxon:10090"  
/clone\_id="A630041P07"  
/tissue\_type="thymus"  
/dev\_stage="3 days neonate"  
/lab\_host="DH10B"  
/clone\_id="RIKEN full-length enriched, 3 days neonate thymus"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGGATTCCTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTAAATACCCGCCCGCC 3']. cDNA was cleaved with XbaI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PUC I."

## ORIGIN

Query Match 100.0%; Score 22; DB 10; Length 458;  
Best Local Similarity 100.0%; Pred. No. 48+02; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0;

Qy 1 AAGTGGGTTCGAGGCTGC 22  
Db 10 AAGTGGGTTCGAGGCTGC 31

## RESULT 3

BF330450/c mRNA 583 bp linear EST 22-NOV-2000  
MR2-PN0364-280800-005-f07 BN0364 Homo sapiens cDNA, mRNA sequence.

## RESULT 4

BF330450 LOCUS BY34796 mRNA 667 bp linear EST 17-DEC-2002  
DEFINITION BY734796 RIKEN full-length enriched, mammary gland RCB-0526  
JY9-MC (A) cDNA Mus musculus cDNA clone G830026K01 5', mRNA  
sequence.

## REFERENCE

## AUTHORS

## KEYWORDS

## SOURCE

## ORGANISM

## ACCESSTION

## VERSION

## KEYWORDS

## EST.

## SOURCE

## ORGANISM

## MUS musculus (house mouse)

## Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## 1 (bases 1 to 667)

## Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

## Nikaido, I., Osoyo N., Saito R., Suzuki H., Yamanaka I.,

## Kiyosawa H., Yagi K., Tomaru Y., Nagami A.,

## Schonbach, C., Gojobori T., Baldarelli R., Hill D.P., Bult C.,

## Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H.,

## Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V.,

## Chothia, C., Corbani L.E., Cousins S., Dalla E., Dragani T.A.,

## Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T.,

## Garibaldi M., Gissi C., Godzik A., Gough J., Grimesmond S.,

## Gustincich S., Hirokawa N., Jackson T.J., Jarvis E.D., Kawai A.,

## Kawai H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A.,

## Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T.,

## Numata K., Okido T., Pavani W.J., Pertea G., Pesole G.,

## Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,

## Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,

## Sandelin A., Schneider C., Semple C.A., Setou M., Shinoda K.,

## Sultana R., Takenaka Y., Teasdale R.D., Tomita M.,

## Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y.,

## Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I.,

## Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P.,

## Hayatsu N., Hirozawa Kishikawa T., Konno H., Nakamura M.,

## Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K.,

## Arakawa T., Fukuda S., Hara A., Hashizume W., Imatori K., Ishii Y.,

## Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K.,

## Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S.,

## Rogers J., Birney E., and Hayashizaki Y.

## Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

tissue mRNA and cDNA amplification were performed under low stringency conditions."

## COMMENT

## Laboratory of Cancer Genetics

## Ludwig Institute for Cancer Research

## Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, São Paulo-SP,

## Brazil

## Tel: +55-11-27049222

## Fax: +55-11-2707003

## Email: asimpson@ludwig.org.br

## This sequence was derived from the FAPESP/LICR Human Cancer Genome

## Project. This entry can be seen in the following URL

## (http://www.ludwig.org.br/scripts/gethtml2.pl?tl1=MR2&amp;t2=MR2- BN0364-

## 280800-005-f07rt3-200-08-28&amp;t4=1)

## Seq. primer: puc 18 forward

## High quality sequence start: 10

## High quality sequence stop: 581.

## Location/Qualifiers

## 1..583 /organism="Homo sapiens"

## /mol "type="mRNA"

## /db\_xref="taxon:9606"

## /dev\_stage="Adult"

## /clone\_id="BN0364"

## /note="Organ: breast normal; Vector: pUC18; Site: 1; Small"

## Site 2: Small; A mini-library was made by cloning products

## derived from ORBESSES PCR (U.S. Letters Patent application

## No. 196,716 - Ludwig Institute for Cancer Research)

## profiles into the pUC 18 vector. Reverse transcription of

## FEATURES

## source

## Contact: Yoshihide Hayashizaki

## Laboratory for Genome Exploration Research Group, RIKEN Genomic

## Sciences Center (GSC), Yokohama Institute

## The Institute of Physical and Chemical Research (RIKEN)

## 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

## Fax: 81-45-503-9216

## URL: http://genome.gsc.riken.go.jp/

## Adachi J., Aizawa K., Akimura T., Arakawa T., Carrinci P.,

## Fukuda S., Hashizume W., Hayashizaki K., Hirozane T., Hori F.,

## Imatori K., Ishii Y., Itoh M., Kagawa J., Kawai J., Kojima Y.,

## Kondo S., Konno H., Koya S., Miyazaki A., Murata M., Nakamura M.,

## Numura K., Numazaki R., Ohno M., Ohsato N., Saito R., Sakazume N.,

Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanuki, A., Muramatsu, M. and Hayashitaki, Y.  
 Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

## source

/organism="Mus musculus"  
 /mol type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="G33026X01"  
 /tissue type="mammary gland"  
 /cell\_line="RCB-0526 JYg-MC(A)"  
 /clone\_id="RIKEN full-length enriched, mammary gland RCB-0526 JYg-MC(A) cDNA"

## ORIGIN

Query Match Score 22; DB 13; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 22; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;  
 Qy 1 AAGGTGGAGTTGGAGGACTGC 22  
 Db 560 AAGGTGGAGTTGGAGGACTGC 581

## RESULT 5

BUT06646 BUT06646 702 bp mRNA linear EST 15-JUL-2003  
 UI-M-FOO-cad-p-11-0-UI.r1 NIH\_BMAP\_FOO Mus musculus CDNA clone  
 DEFINITION IMAGE:6409978\_5', mRNA sequence.  
 ACCESSION BUT06646  
 VERSION BUT06646.1 GI:23637322  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1-(bases 1 to 702)  
 AUTHORS NIH-MGC http://mgc.ncbi.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgsbs-r@mail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 DNA Distribution by: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

## FEATURES

## source

seq primer: PYX-5.  
 Location/Qualifiers 1. .702  
 /organism="Mus musculus"  
 /mol type="mRNA"  
 /db\_xref="taxon:8265"  
 /clone\_id="sp042ks"  
 /sex="female"  
 /dev\_stage="adult"  
 /clone\_id="Winter flounder spleen"  
 /note="Organ: spleen"

/mol type="mRNA"  
 /strain="P57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6409978"  
 /tissue type="whole brain"  
 /dev\_stage="embryo 12.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_id="NIH\_BMAP\_FOO"  
 /note="Organ: Brain; Vector: PYX-ASC; Site: 1; EcoRI I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoRI adaptor, digested with NotI and then cloned directionally into PYX-ASC vector. The library tag sequence located between the Not I site and the polyA tail is TGAGAGGCC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

Query Match Score 22; DB 13; Length 702;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 22; Conservative 0; Mismatches 0;  
 Indels 0; Gaps 0;  
 Qy 1 AAGGTGGAGTTGGAGGACTGC 22  
 Db 314 AAGGTGGAGTTGGAGGACTGC 335

RESULT 6  
 AW013379 LOCUS sp042ks winter flounder spleen Pseudopleuronectes americanus cDNA  
 DEFINITION AW013379 clone sp042ks 5', similar to C53B4.4 (Genorhabditis elegans), mRNA sequence.  
 ACCESSION AW013379.1 GI:5862157  
 VERSION EST.  
 KEYWORDS Pseudopleuronectes americanus (winter flounder)  
 SOURCE ORGANISM Pseudopleuronectes americanus  
 Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygi; Teleostei; Euteleoste; Neoteleoste;  
 Acanthomorpha; Acanthopterygii; Percmonorpha; Pleuronectiformes; Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.  
 REFERENCE 1 (bases 1 to 723)  
 AUTHORS Douglas,S.E., Gallant,J.W., Bullerwell,C.E., Wolff,C., Munholland,J. and Reith,M.E.  
 TITLE Winter flounder expressed sequence tags: Establishment of an EST database and identification of novel fish genes  
 JOURNAL Marine Biotechnology (1999) In press  
 COMMENT Contact: Reith M  
 Marine Biology  
 NRC Institute for Marine Biosciences  
 1411 Oxford St., Halifax, Nova Scotia, B3H 3Z1, Canada  
 Tel: (902) 426-8276  
 Fax: (902) 426-9413  
 Email: michael.reith@nrc.ca  
 Seq Primer: M13 Forward.  
 Location/Qualifiers 1..723  
 /organism="Pseudopleuronectes americanus"  
 /mol type="mRNA"  
 /db\_xref="taxon:8265"  
 /clone\_id="sp042ks"  
 /sex="female"  
 /dev\_stage="adult"  
 /clone\_id="Winter flounder spleen"  
 /note="Organ: spleen"



BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr Web : <a href="http://www.genoscope.cns.fr/">www.genoscope.cns.fr</a> Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5483.r. For more information about this cluster, see <a href="http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS2BAX162B10_AX272D11_1&amp;cluster=5483.r">http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS2BAX162B10_AX272D11_1</a> .	PUBMED REFERENCE 4 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. TITLE JOURNAL JOURNAL REFERENCE 5 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. TITLE JOURNAL JOURNAL REFERENCE 6 AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furano,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hirakawa,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kasukawa,T., Katoch,H., Kawai,J., Koijima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohzato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibusawa,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Takanashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
FEATURES source 1. Location/Qualifiers 1..944 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DK002YA12" /cell_type="HELA CELLS COT 25-NORMALIZED" /cell_line="HELA" /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the PCMVSPORT 6 vector. Library was normalized."	JOURNAL REFERENCE 1 AUTHORS Direct Submission Submitted (16-JUN-2001). Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 23-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, Fax:81-45-503-9222, URL: <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> , Tel:81-45-503-9222)
ORIGIN Query Match Best Local Similarity 100.0% ; Score 22; DB 13; Length 944; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	JOURNAL REFERENCE 2 COMMENT CDNA 1 library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: <a href="http://phantom.gsc.riken.go.jp/">http://phantom.gsc.riken.go.jp/</a> . COMMENT CDNA 1 library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: <a href="http://phantom.gsc.riken.go.jp/">http://phantom.gsc.riken.go.jp/</a> .
RESULT 10 AK041847 LCUS DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A63n041P07 product-hypothetical PDZ domain (also known as DHR or GLGF) containing protein, full insert sequence.	FEATURES Source 1..2052 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:A63n041P07" /db_xref="MGI:2406859" /clone="A63n041P07" /tissue_type="thymus" /clone_id="RIKEN Full-length enriched mouse cDNA library" /dev_stage="3 days neonate" /note="hypothetical PDZ domain (also known as DHR or GLGF) containing protein (InterPro IPR001478, evidence: InterPro)" ORIGIN
ACCESSION AK041847 VERSION AK041847.1 KEYWORDS HTC; AP trapper. SOURCE Mus musculus (house mouse) ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	REFERENCE 1 AUTHORS Carninci,P. and Hayashizaki,Y. TITLE High-efficiency full-length cDNA cloning JOURNAL Meth. Enzymol. 303, 19-44 (1999) MEDLINE 9927253 PUBLISHED 10349636
REFERENCE 2 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. TITLE Normalization and subtraction of cap-trapper selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes JOURNAL Genome Res. 10 (10), 1617-1630 (2000) MEDLINE 20493374 PUBLMED 11042159	REFERENCE 3 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Niishima,N., Iishi,Y., Nakamura,S., Hazama,M., Nishimura,T., Harada,I., Yamamoto,R., Matsumoto,H., Saiguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanuki,M., Okazaki,Y., Ishikawa,T., Ozawa,K., Matsuo,S., Kawai,J., Yoneda,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y. TITLE RIKEN integrated sequence analysis (RISA) system--384 multicapillary sequencer JOURNAL Genome Res. 10 (11), 1757-1771 (2000) MEDLINE 20530913
RESULT 11	

BC033019	LOCUS	BC033019	3069 bp mRNA, linear	HTC 19-NOV-2003	Query Match 100.0%; Score 22; DB 11;
DEFINITION	Homo sapiens hypothetical protein LOC11987, mRNA (cDNA clone IMAGE:4821752), containing frame-shift errors.			Best Local Similarity 100.0%; Pred. No. 76+02;	Mismatches 0; Gaps 0;
ACCESSION	BO033019				
VERSION	BC033019.1	GI:21542573			
KEYWORDS	HTC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shemmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Sores, M.B., Bandal, M.P., Asavart, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Tsohomyoki, S., Carninci, P., Prange, C., Raha, S. S., Loqueland, N.A., Peters, G.J., Abramson, R.D., Mulay, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Pahay, J., Heitton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schein, J.E., Jones, S.J. and Marra, M.A.				
AUTHORS					
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
MEDLINE	12388957				
PUBLMED	12477932				
REFERENCE	2 (bases 1 to 3069)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: <a href="http://mgc.ncbi.nih.gov">http://mgc.ncbi.nih.gov</a>				
COMMENT	Contact: MGC help desk Email: <a href="mailto:crgobs@nlm.nih.gov">crgobs@nlm.nih.gov</a>				
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.				
	CDNA Library Preparation: Michael J. Brownstein (NHRGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)				
	CDNA Library Arrayed By: The I.M.A.G.E. Consortium (LILNL)				
	DNA Sequencing by: Institute for Systems Biology <a href="http://www.systemsbiology.org">http://www.systemsbiology.org</a>				
	Contact: amadasysystemsbiology.org				
	Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting				
FEATURES	source				
	1 3069				
	/organism="Homo sapiens"				
	/mol type="mRNA"				
	/db_xref="Taxon:9606"				
	/clone="IMAGE:4821752"				
	/tissue type="Testis"				
	/clone_id="NIH MGC_97"				
	/lab_host="DH10B"				
	/note="Vector: pBluescript"				
ORIGIN					

CRL-1751 WEHI 164#  
/clone\_id="RIKEN full-length enriched, CRL-1751 WEHI 164#  
cDNA"



KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 989)  
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization.  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Séquençage  
BP 191 91006 EVRY Cedex - France  
Email: seqref@genoscope.cns.fr. Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5483.r. For more information about this cluster, see <http://www.Genoscope.fr/cgi-bin/cluster.cgi?seq=CS0DK002BA06QPI&cluster=5483.r>. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DK002BA06QPI.  
FEATURES Location/Qualifiers  
source 1. .989  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DK002BA12"  
/cell\_type="HELA CELLS COT 25-NORMALIZED"  
/cell\_line="HELA"  
/clone\_id="Homo sapiens HEla CELLS COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."  
ORIGIN

Query Match 90.9%; Score 20; DB 13; Length 989;  
Best Local Similarity 90.9%; Pred. No. 2.8e+03;  
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAGGTGGACTCGAGGAGCTGC 22  
Db 637 AARGTGGAAATCGAGGAGTC 658

Search completed: July 18, 2004, 14:36:35  
Job time : 275.872 secs



Run on:	July 18, 2004, 11:31:29 ; Search time 36.7148 Seconds (without alignments)			
OM nucleic - nucleic search, using sw model	2545.580 Million cell updates/sec			
Title:	US-10-051-769-6			
Perfect score:	22			
Sequence:	1 gtggaaaggccgcgttgtactcc 22			
Scoring table:	IDENTITY_NUC			
Gapop:	10.0 , Gapext 1.0			
Searched:	3373853 seqs, 2124099041 residues			
Total number of hits satisfying chosen parameters:	6747726			
Minimum DB seq length:	0			
Maximum DB seq length:	2000000000			
Post-processing:	Minimum Match 0% Maximum Match 100% Listing first 45 summaries			
Database :	N_Geneseq_29Jan04:*			
	1: geneseqn1981s:*			
	2: geneseqn1991s:*			
	3: geneseqn2001s:*			
	4: geneseqn2004s:*			
	5: geneseqn2005s:*			
	6: geneseqn2006s:*			
	7: geneseqn2007s:*			
	8: geneseqn2008s:*			
	9: geneseqn2009s:*			
	10: geneseqn2014s:*			
Pred.	No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES	*			
Result No.	Score	Query Match Length	DB ID	Description
1	22	100.0	22	ABX25683 Human Gli
2	22	100.0	22	AAD61563 Human Gli
c 3	22	100.0	263	ABX25681 Rat Gli
c 4	22	100.0	263	AAD61561 Human Gli
c 5	20.4	92.7	261	ABX25679 Human Gli
c 6	20.4	92.7	261	AAD61560 Human Gli
c 7	20.4	92.7	873	ABX25680 Human Gli
c 8	20.4	92.7	960	ABX25678 Human Gli
c 9	20.4	92.7	960	AAD61559 Human Gli
c 10	20.4	92.7	1092	ADE07789 Novel cod
c 11	20.4	92.7	2639	ADB2530 Human Gli
c 12	20.4	92.7	3832	AAD61564 Human Gli
c 13	20.4	92.7	3847	ABX27764 Human cDNA
c 14	20.4	92.7	4450	AAK1608 Human pol
c 15	20.4	92.7	4470	ABA09174 Human sec
c 16	20.4	92.7	4470	AAK2592 Human pol
c 17	20.4	92.7	4470	ADE09891 Novel DNA
c 18	20.4	92.7	4801	ADD9052 Human nov
c 19	18.8	85.5	879	AAQ14436 Mutant Cyt
c 20	18.8	85.5	879	AAQ24437 Mutant m
c 21	18.8	85.5	960	AAQ36210 DNA encod
c 22	17.4	79.1	1935	ABZT76503 Lactobaci
c 23	17.4	79.1	8056	ABZT76501 Lactobaci
				ABZT76576 Lactobaci
				Aat46934 Human von WVF GPIB
				Aaq13428 VWF GPIB
				Aat78689 VWF GPIB
				Aar08900 Von Wille
				Aav14157 Human mat
				Aaz56182 Plasmid p
				Aaq46008 cDNA enco
				Aaq14803 Von Wille
				Aaq18233 von Wille
				Aaa15183 DNA encod
				Aah1440 Human cDN
				Aad56131 Human NOT
				Ada02494 Human NOT
				Adb7232 Human NOT
				Aad45961 HSA_vWF(4
				Aad30999 Notch cDNA
				Aal15428 Human Not
				Aav08901 Von Wille
				Aav74156 Human mat
				Aaz56177 Human von Wille
				Abi166946 Lung canc

CC sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into CC Glioblastoma Multiforme, and a kit for use in detecting whether a patient is at risk for progression into Glioblastoma Multiforme comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into Glioblastoma Multiforme (a type of brain cancer). The EST useful as molecular marker for tumour cell identification and for diagnosing or identifying candidates at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is a human GLITEN gene PCR primer

XX Sequence 22 BP; 3 A; 7 C; 7 G; 5 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 22; DB 7; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.98; O; Mismatches 0; Indels 0; Gaps 0;

Db 1 GTGGAAGGCCGCGTGTACTCC 22

Db 1 GTGGAAGGCCGCGTGTACTCC 22

RESUL 2 AD61563 standard; DNA; 22 BP.

XX AC AD61563;

XX DT 15-JAN-2004 (first entry)

DR Human GLITEN cDNA amplifying PCR primer #2.

XX Human; Glioblastoma multiforme; GBM; GLITEN; brain cancer; diagnosis; gene therapy; PCR; primer; ss.

XX OS Homo sapiens.

XX US2003108915-A1.

XX PN 20-OCT-2000; 2000US-0242160P.

XX PR 20-OCT-2001; 2001US-00051769.

XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX PT 20-AUG-2002; 2002US-00224624.

XX SQ 20-OCT-2000; 2000US-0242160P.

XX DR 16; Page 5; Opp; English.

XX PS Novel Glioblastoma multiforme associated protein GLITEN useful for treating Glioblastoma multiforme and diagnosing brain cancer.

XX CC The present invention provides novel Glioblastoma multiforme (GBM) associated protein GLITEN useful for treating glioblastoma multiforme and diagnosing brain cancer. The invention is useful in brain cancer therapy, treatment and diagnosis. The invention is also useful in gene therapy.

XX CC The present sequence is human GLITEN cDNA amplifying PCR primer

XX CC The present invention provides novel Glioblastoma multiforme (GBM) associated protein GLITEN useful for treating glioblastoma multiforme and diagnosing brain cancer. The invention is useful in brain cancer therapy, treatment and diagnosis. The invention is also useful in gene therapy.

XX CC The present sequence is a Rat GLITEN expressed sequence tag

XX SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

XX DR WPI; 2003-810848/76.

XX PT Novel Glioblastoma multiforme associated protein GLITEN useful for treating Glioblastoma multiforme and diagnosing brain cancer.

XX PT The present invention provides novel Glioblastoma multiforme (GBM) associated protein GLITEN useful for treating glioblastoma multiforme and diagnosing brain cancer. The invention is useful in brain cancer therapy, treatment and diagnosis. The invention is also useful in gene therapy.

XX CC The present sequence is human GLITEN cDNA amplifying PCR primer

XX SQ Sequence 22 BP; 3 A; 7 C; 7 G; 5 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 22; DB 9; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.1; O; Mismatches 0; Indels 0; Gaps 0;

Db 1 GTGGAAGGCCGCGTGTACTCC 22

Db 244 GTGGAAGGCCGCGTGTACTCC 22

Qy 1 GTGGAAGGCCGCGTGTACTCC 22

Qy 1 GTGGAAGGCCGCGTGTACTCC 22

CC	Db	1 GTGGAAGGCCGCGTGTACTCC 22
CC		RESULT 3 ABX95681 /C ID ABX95681 standard; cDNA; 263 BP. XX
CC		AC ABX95681; XX
CC		DT 30-JUN-2003 (first entry) XX
CC		DE Rat GLITEN expressed sequence tag. XX
CC		KW Rat; ss; EST; glioblastoma multiforme; brain cancer; GLITEN; GMB; KW expressed sequence tag. XX
CC		OS Rattus Norvegicus. XX
CC		PN US2003044811-A1. XX
CC		PD 06-MAR-2003. XX
CC		PP 20-OCT-2001; 2001US-00051769. XX
CC		PR 20-OCT-2000; 2000US-0242160P. XX
CC		PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY. XX
CC		PI McKinnon RD; XX
CC		DR WPI; 2003-418995/39. XX
CC		PT New isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLITEN, useful as a molecular marker for tumor cell identification and classification. XX
CC		PS Example 2; Page 4; 11pp; English. XX
CC		The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLITEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into Glioblastoma Multiforme), and a kit for use in detecting whether a patient is at risk for progression into Glioblastoma Multiforme comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into Glioblastoma Multiforme (a type of brain cancer). The EST useful as molecular marker for tumour cell identification and classification at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is a Rat GLITEN expressed sequence tag XX
CC		SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other; Best Local Similarity 100.0%; Pred. No. 1.1; O; Mismatches 0; Indels 0; Gaps 0;
CC		Query Match 100.0%; Score 22; DB 7; Length 263; Matches 22; Conservative 0; O; Mismatches 0; Indels 0; Gaps 0;
CC		Qy 1 GTGGAAGGCCGCGTGTACTCC 22
CC		Db 244 GTGGAAGGCCGCGTGTACTCC 22

**RESULT 4**  
 AAD61561/c  
 ID AAD61561 standard; cDNA; 263 BP.  
 XX  
 AC AAD61561;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX Human Gliten EST clone cDNA #2.  
 DE Human; glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;  
 KW gene therapy; expressed sequence tag; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003108915-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PT 20-AUG-2002; 2002US-00224624.  
 PR 20-OCT-2000; 2000US-0242160P.  
 PR 20-OCT-2001; 2001US-00051769.

(UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

McKinnon RD;

XX  
 DR WPI; 2003-418995/39.

XX New isolated nucleic acid representing a gene product associated with  
 glioblastoma Multiforme, designated as Gliten, useful as a molecular  
 marker for tumor cell identification and classification.  
 XX  
 PS Claim 1; Page 7; 11pp; English.

XX  
 CC The invention relates to an isolated nucleic acid representing a gene  
 product associated with Glioblastoma Multiforme, designated as Gliten,  
 comprising the human EST (expressed sequence tag) appearing as ABX95678,  
 or a sequence that hybridises under stringent conditions to the EST, or  
 its complement. Also included are a probe for use in identifying a  
 patient at risk for progression into the malignant phenotype, comprising  
 the nucleic acids detailed above, detecting whether a patient is at risk  
 for progression into Glioblastoma Multiforme (comprising: (a) providing a  
 sample from a patient; (b) adding the probe to the sample or performing  
 PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)  
 analysing levels of mRNA bound with the probe; and (d) treating a control  
 sample to assess the level of mRNA presence in the control sample, where the  
 presence of increased levels of mRNA expression in the sample in an  
 amount higher than the control sample indicates risk for progression into  
 Glioblastoma Multiforme), and a kit for use in detecting whether a  
 patient is at risk for progression into Glioblastoma Multiforme  
 (comprising the probe or primers). The probe or the kit is useful for  
 detecting whether a patient is at risk for progression into Glioblastoma  
 Multiforme (a type of brain cancer). The EST useful as molecular marker  
 for tumour cell identification and classification and for diagnosing or  
 identifying candidates at risk for progression into a malignant phenotype  
 especially in brain cancer therapy, treatment and diagnosis. The present  
 sequence is the human Gliten expressed sequence tag

XX  
 SQ Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;  
 XX  
 Query Match 92.7%; Score 20.4%; DB 7; Length 261;  
 Best Local Similarity 95.5%; Pred. No. 6.4%;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 QY 1 GTGAAAGCCGCCGTGTACTCC 22  
 Db 244 GTGAAAGCCGCCGTGTACTCC 223

**RESULT 5**  
 ABX95679/c  
 ID ABX95679 standard; cDNA; 261 BP.  
 XX  
 AC ABX95679;  
 XX  
 DT 30-JUN-2003 (first entry)  
 XX Human Gliten expressed sequence tag.  
 DE Homo sapiens.  
 XX Human; ss; EST; chromosome 10q25; glioblastoma multiforme; brain cancer;  
 KW Gliten; expressed sequence tag.  
 XX  
 OS US2003044811-A1.  
 XX  
 PN US2003108915-A1.

XX  
 PD 12-JUN-2003.  
 XX  
 PP 20-AUG-2002; 2002US-00224624.  
 XX  
 PR 20-OCT-2000; 2000US-0242160P.  
 PR 20-OCT-2001; 2001US-00051769.

XX  
 PR 20-OCT-2000; 2000US-0242160P.  
 XX  
 PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX  
 PI McKinnon RD;

XX  
 DR WPI; 2003-418995/39.

XX New isolated nucleic acid representing a gene product associated with  
 glioblastoma Multiforme, designated as Gliten, useful as a molecular  
 marker for tumor cell identification and classification.  
 XX  
 PS Claim 1; Page 7; 11pp; English.

XX  
 CC The invention relates to an isolated nucleic acid representing a gene  
 product associated with Glioblastoma Multiforme, designated as Gliten,  
 comprising the human EST (expressed sequence tag) appearing as ABX95678,  
 or a sequence that hybridises under stringent conditions to the EST, or  
 its complement. Also included are a probe for use in identifying a  
 patient at risk for progression into the malignant phenotype, comprising  
 the nucleic acids detailed above, detecting whether a patient is at risk  
 for progression into Glioblastoma Multiforme (comprising: (a) providing a  
 sample from a patient; (b) adding the probe to the sample or performing  
 PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)  
 analysing levels of mRNA bound with the probe; and (d) treating a control  
 sample to assess the level of mRNA presence in the control sample, where the  
 presence of increased levels of mRNA expression in the sample in an  
 amount higher than the control sample indicates risk for progression into  
 Glioblastoma Multiforme), and a kit for use in detecting whether a  
 patient is at risk for progression into Glioblastoma Multiforme  
 (comprising the probe or primers). The probe or the kit is useful for  
 detecting whether a patient is at risk for progression into Glioblastoma  
 Multiforme (a type of brain cancer). The EST useful as molecular marker  
 for tumour cell identification and classification and for diagnosing or  
 identifying candidates at risk for progression into a malignant phenotype  
 especially in brain cancer therapy, treatment and diagnosis. The present  
 sequence is the human Gliten expressed sequence tag

XX  
 SQ Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;  
 XX  
 Query Match 92.7%; Score 20.4%; DB 7; Length 261;  
 Best Local Similarity 95.5%; Pred. No. 6.4%;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 QY 1 GTGAAAGCCGCCGTGTACTCC 22  
 Db 244 GTGAAAGCCGCCGTGTACTCC 223

**RESULT 6**  
 AAD1560/c  
 ID AAD1560 standard; cDNA; 261 BP.  
 XX  
 AC AAD1560;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Human Gliten EST clone cDNA #1.  
 XX  
 KW Human; glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;  
 KW gene therapy; expressed sequence tag; ss.  
 XX  
 OS Homo sapiens.

XX  
 PN US2003108915-A1.

XX  
 PD 12-JUN-2003.

XX  
 PP 20-AUG-2002; 2002US-00224624.

XX  
 PR 20-OCT-2000; 2000US-0242160P.

PR 20-OCT-2001; 2001US-00051769.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 PA  
 XX  
 PI McKinnon RD;  
 XX  
 DR WPI; 2003-810848/76.

XX Novel glioblastoma multiforme associated protein Gliten useful for  
 PT treating glioblastoma multiforme and diagnosing brain cancer.  
 XX  
 PS Claim 38; Page 11; Opp; English.

XX The present invention provides novel glioblastoma multiforme (GBM)  
 CC associated protein Gliten useful for treating glioblastoma multiforme and  
 CC diagnosing brain cancer. The invention is useful in brain cancer therapy,  
 CC treatment and diagnosis. The invention is also useful in gene therapy.  
 CC The present sequence is human Gliten EST clone cDNA. This EST clone is  
 CC used in the exemplification of the invention.

XX Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

SQ Query Match 92.7%; Score 20.4%; DB 9; Length 261;  
 Best Local Similarity 95.5%; Pred. Nc. 6.4%;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 244 GTGGAAGCCGCCGTGACTCC 22  
 1 GTCGAAAGCCGCCGTGACTCC 22  
 244 GTGGAAGCCGCCGTGACTCC 223

RESULT 7  
 ABX95610/c  
 ID ABX95610 standard; DNA; 873 BP.  
 AC ABX95610;  
 XX  
 DT 30-JUN-2003 (first entry)

XX Human Gliten open reading frame.  
 KW Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;  
 KW Gliten; GMB.  
 XX  
 OS Homo sapiens.  
 XX  
 Key Location/Qualifiers  
 CDS 1..873  
 FT /tag= a  
 FT /partial  
 FT /product= "GLITEN"  
 FT /note= "No stop codon shown."  
 XX US2003044811-A1.  
 PN  
 PD 06-MAR-2003.  
 XX  
 PF 20-OCT-2001; 2001US-00051769.  
 XX  
 PR 20-OCT-2000; 2000US-0242160P.  
 XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 XX  
 PI McKinnon RD;  
 XX  
 DR WPI; 2003-418995/39.

XX New isolated nucleic acid representing a gene product associated with  
 PT Glioblastoma Multiforme, designated as Gliten, useful as a molecular  
 PT marker for tumor cell identification and classification.  
 XX  
 Disclosure; Page 7; 11pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid representing a gene

CC product associated with Glioblastoma Multiforme, designated as Gliten,  
 CC comprising the human EST (expressed sequence tag) appearing as ABX95678,  
 CC or a sequence that hybridises under stringent conditions to the EST, or  
 CC its complement. Also included are a probe for use in identifying a  
 CC patient at risk for progression into the malignant phenotype, comprising  
 CC the nucleic acids detailed above, detecting whether a patient is at risk  
 CC for progression into Glioblastoma Multiforme (comprising: (a) providing a  
 CC sample from a patient; (b) adding the probe to the sample or performing  
 CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)  
 CC analysing levels of mRNA bound with the probe; and (d) treating a control  
 CC sample to assess the level of mRNA in a control sample, where the  
 CC presence of increased levels of mRNA expression in the sample in an  
 CC amount higher than the control sample indicates risk for progression into  
 CC Glioblastoma Multiforme), and a kit for use in detecting whether a  
 CC patient is at risk for progression into Glioblastoma Multiforme  
 CC (comprising the probe or primers). The probe or the kit is useful for  
 CC detecting whether a patient is at risk for progression into Glioblastoma  
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker  
 CC for tumour cell identification and classification and for diagnosing or  
 CC identifying candidates at risk for progression into a malignant phenotype  
 CC especially in brain cancer therapy, treatment and diagnosis. The present  
 CC sequence is the open reading frame from the Gliten gene

Query Match 92.7%; Score 20.4%; DB 7; Length 873;  
 Best Local Similarity 95.5%; Pred. No. 6.8%;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGAAAGCCGCCGTGACTCC 22  
 Db 609 GTGGAAGCCGCCGTGACTCC 588

RESULT 8  
 ABX95678/c  
 ID ABX95678 standard; DNA; 960 BP.  
 XX  
 AC ABX95678;  
 XX  
 DT 30-JUN-2003 (first entry)  
 XX  
 DE Human gene encoding Gliten.  
 XX  
 KW Human; ds; gene; chromosome 10q25; glioblastoma multiforme; brain cancer;  
 KW Gliten; GMB.  
 XX  
 OS Homo sapiens.  
 XX  
 Key Location/Qualifiers  
 CDS 30..906  
 FT /\*tag= a  
 FT /product= "GLITEN"  
 FT /note= "No stop codon shown."  
 XX US2003044811-A1.  
 PN  
 PD 06-MAR-2003.  
 XX  
 PF 20-OCT-2001; 2001US-00051769.  
 XX  
 PR 20-OCT-2000; 2000US-0242160P.  
 XX  
 PR (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 XX  
 PA McKinnon RD;  
 XX  
 DR WPI; 2003-418995/39.  
 XX  
 PT New isolated nucleic acid representing a gene product associated with  
 PT Glioblastoma Multiforme, designated as Gliten, useful as a molecular  
 PT marker for tumor cell identification and classification.  
 XX  
 Disclosure; Page 7; 11pp; English.  
 XX  
 CC Example 2; Fig 1; 11pp; English.

The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as ABX95678, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into Glioblastoma Multiforme), and a kit for use in detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into Glioblastoma Multiforme (a type of brain cancer). The EST useful as molecular marker for tumour cell identification and classification and for diagnosing or identifying candidates at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is a region of human chromosome 10 (q25) containing the GITEN gene.

RESULT 9  
D61559/c  
AAD61559 standard; cDNA; 960 BP.  
AAD61559;  
15-JAN-2004 (first entry)  
Human Gliten partial cDNA.  
Human glioblastoma multiforme: GBM: Gliten; brain cancer; diagnosis;  
glioma; glioblastoma; glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;

*Homo sapiens.*  
US2003108915-A1.  
12-JUN-2003.

20-OCT-2000; 2000US-0242160P.  
20-OCT-2001; 2001US-0005179.  
(YUNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
McKinnon RD;  
WPI; 2003-81084B/76.

Novel glioblastoma multiforme associated protein Gliten useful for treating glioblastoma multiforme and associated compositions

Example 2: Fig 1; opp; English.  
The present invention provides novel combinations comprising

CC associated protein Gliten useful for treating glioblastoma multiforme and  
 CC diagnosing brain cancer. The invention is useful in brain cancer therapy,  
 CC treatment and diagnosis. The invention is also useful in gene therapy.  
 CC The present sequence is human Glitrn partial cDNA. The human Glitrn gene  
 CC is located on chromosome 10. This partial cDNA is used in the  
 CC exemplification of the invention.

XX

SQ	Sequence 960 BP;	146 A;	324 C;	329 G;	161 T;	0 U;	0 Other;
Query Match	92.7%	Score 20.4;	DB 9;	Length 960;			
Best Local Similarity	95.5%;	Pred. No. 6.9;					
Matches 21;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps 0;
Qy	1	GTGGAAGCCGGCGTGTACTCC	22				
Dn	639	GTGGAAGCCGGCGTGTACTCC	619				

RESULT 10  
ADE07789/c  
ID ADE07789 standard; DNA; 1092 BP.  
XX AC ADE07789;  
XX DT 29-JAN-2004 (first entry)

DE Novel coding sequence (useful for identifying genetic disorders) #855.  
 XX novel gene; novel protein; tissue marker; molecular weight marker;  
 KW chromosome marker; genetic disorder; gene; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN WO2003054152-A2.

PD 03-JUL-2003 .  
XX PF 10-DEC-2002; 2002WO-US039555.  
XX PR 10-DEC-2001; 2001US-0339739P.  
PR 11-DEC-2001; 2001US-0339453P.  
PR 14-MAR-2002; 2002US-0365591P.  
PR 14-MAR-2002; 2002US-0365384P.

PR 1.2-ARR-2002; 2002US-U374281P.  
 PR 22-ARR-2002; 2002US-03742815P.  
 PR 24-ARR-2002; 2002US-00128558.  
 PR 24-ARR-2002; 2002US-0376045P.  
 XX (HYSEQ INC.  
 PA Tang YT, Asundi V, Goodrich RW,  
 PI Ren F, Zhang J, Zhao QA, Wang J;

PI	Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX	
DR	WPI; 2003-569235/53.
DR	P-PSBB; ADE08700.
XX	
PT	New polynucleotides, useful for expressing recombinant proteins for analysis, characterization or therapeutic use, or as markers for tissues
PT	

XX Claim 1; SEQ ID NO 855; 1177pp; English.  
PS  
XX The invention comprises the amino acid and coding sequences of novel  
CC proteins. The DNA and protein sequences of the invention are useful as  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present DNA sequence represents a gene of the  
CC invention.  
XX

Query Match 92.7%; Score 20.4; DB 9; Length 1092;  
 Best Local Similarity 95.5%; Pred. No. 6.9;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCCCTGTACTCC 22  
 Db 609 CTGGAAAGCCCCGGTGTACTCC 588

---

RESULT 11  
 ADB62330/c  
 ID ADB62330 standard; cDNA; 2639 BP.  
 XX  
 AC ADB62330;  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human cDNA encoding clone HHDPc20088160.  
 XX  
 KW Human; SS; Gene; pharmaceutical; diagnostic; gene therapy;  
 KW tissue regeneration; cell regeneration; membrane protein;  
 KW signal transduction-related protein; transcription-related protein;  
 KW osteoporosis; neurological disease; cancer; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT CDS  
 FT /  
 FT /\*tag= a  
 FT /partial  
 FT /product= "Clone HHDPc20088160 protein"  
 XX  
 PN EP1308459-A2.  
 XX  
 PD 07-MAY-2003.  
 XX  
 PF 28-MAR-2002; 2002EP-00007401.  
 XX  
 PR 05-NOV-2001; 2001JP-00379298.  
 XX  
 PA (HELI-) HELIX RES INST.  
 (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
 Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
 Seki N, Yoshikawa T, Otsuka M, Nagahari K, Matsuo Y,  
 WPI; 2003-450961/43.  
 DR P-PSDB; ADB64500.  
 XX  
 PT New polynucleotides and polypeptides, useful for developing a diagnostic  
 marker or medicines for regulation of their expression and activity, or  
 as targets of gene therapy.  
 PS Claim 1: Page; 222pp; English.

XX  
 CC The invention discloses a polynucleotide comprising a sequence selected  
 from 1970 fully defined nucleotide sequences which encode novel  
 polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 or its partial peptide, an antibody binding to the polypeptide or peptide  
 of the polynucleotide, immunologically assaying the polypeptide or  
 peptide of the polynucleotide by contacting the polypeptide or peptide  
 with the antibody of the encoded protein, and observing the binding  
 between the two, a transformant carrying the polynucleotide in an  
 expressible manner and an antisense polynucleotide. The oligonucleotide  
 is useful as a primer for synthesising the polynucleotide, or as a probe  
 for detecting the polynucleotide. The polynucleotides and encoded  
 proteins are useful as pharmaceutical agents and many disease-related  
 genes may be included in them, for developing a diagnostic marker or  
 medicines for regulation of their expression and activity, or as targets  
 of gene therapy. The genes are involved in tissue and/or cell

CC regeneration. Membrane proteins, signal-transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
 CC data for this patent is not represented in the printed specification, but  
 CC is based on sequence information supplied by the European Patent Office.  
 XX  
 SQ Sequence 2639 BP; 657 A; 702 C; 719 G; 561 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 9; Length 2639;  
 Best Local Similarity 95.5%; Pred. No. 7.3;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGAAGCCGCCCTGTACTCC 22  
 Db 822 GTGAAAGCCCCGGTGTACTCC 801

---

RESULT 12  
 AAD6154/c  
 ID AAD61564 standard; cDNA; 3832 BP.  
 XX  
 AC AAD61564;  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Human Gliten cDNA.  
 XX  
 KW Human; glioblastoma multiforme; GBM; GLITEN; brain cancer; diagnosis;  
 KW gene therapy; chromosome 10; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT CDS  
 FT /  
 FT /\*tag= a  
 /product= "Human GLITEN protein"  
 /note= "The CDS is specifically claimed in claim 1"  
 FT primer\_bind  
 FT /  
 FT /\*tag= b  
 /bound\_moiety= "5' primer #843"  
 FT primer\_bind  
 FT /  
 FT /\*tag= c  
 /bound\_moiety= "5' primer #1405"  
 FT primer\_bind  
 FT /  
 FT /\*tag= d  
 /bound\_moiety= "5' primer #2083"  
 FT primer\_bind  
 FT /  
 FT /\*tag= e  
 /bound\_moiety= "5' primer #2335"  
 FT primer\_bind  
 FT /  
 FT /\*tag= f  
 /bound\_moiety= "3' primer #3636"  
 XX  
 PN US2003108915-A1.  
 XX  
 PD 12-JUN-2003.  
 XX  
 PF 20-AUG-2002; 2002US-00224624.  
 XX  
 PR 20-OCT-2000; 2000US-0242160P.  
 XX  
 PA 20-OCT-2001; 2001US-00051769.  
 XX  
 PA (UYN-E) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 XX  
 PI McKinnon RD;  
 XX  
 DR WPI; 2003-10848/76.  
 DR P-PSDB; ABN00758.  
 XX  
 PT Novel glioblastoma multiforme associated protein GLITEN useful for

PT treating glioblastoma multiforme and diagnosing brain cancer.  
 XX Claim 1; Fig 4; Opp; English.

PS The present invention provides novel glioblastoma multiforme (GBM) associated protein Gliten useful for treating glioblastoma multiforme and diagnosing brain cancer. The invention is useful in brain cancer therapy, treatment and diagnosis. The invention is also useful in gene therapy. The present sequence is human Gliten cDNA. The human Gliten gene is located on chromosome 10.

XX Sequence 3832 BP; 1090 A; 926 C; 957 G; 859 T; 0 U; 0 Other;  
 Best Local Similarity 95.5%; Pred. No. 7.4;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCCGTTGTACTCC 22  
 Db 786 GTGGAAGCCGCCGTTGTACTCC 765

Query Match Score 20.4; DB 9; Length 3832;  
 SQ Sequence 3847 BP; 1197 A; 744 C; 819 G; 1086 T; 0 U; 1 Other;

RESULT 13 ABK87764/C ID ABK87764 standard; cDNA; 3847 BP.  
 XX DT 24-SEP-2002 (first entry)  
 XX DE Human cDNA encoding glioma antigen KU-GB-5.  
 KW Human; SS; gene; glioma antigen; KU-GB-5; glioma; cytostatic; cancer;  
 KW immunostimulant; immunosuppressive; peptide therapy; gene therapy;  
 KW malignant brain tumour.  
 OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FT 211..2883  
 FT /\*tag= a  
 FT /product= "KU-GB-5"  
 XX WO200255695-A1.  
 PN 18-JUL-2002.  
 PD 09-JAN-2001; 2001WO-00001965.  
 PA (UYKE-) UNIV KEIO.  
 XX PI Toda M, Kawakami Y, Kawase T, Iizuka Y;  
 XX DR WPI; 2002-538141/57.  
 DR P-PSDB; AA199614.  
 XX PT New human glioma antigen for diagnosing and treating Glioma, human malignant brain tumor and other cancers, and for studying onset of Glioma.  
 PT XX Claim 12; Page 94-100; 109pp; Japanese.  
 XX The invention relates to preparing glioma antigen and/or glioma antigen gene comprises e.g. extracting and isolating total RNA from a glioma cell line then synthesising cDNA for constructing a lambda phage cDNA library and reacting the library with the serum for reaction and detecting positive clones reacting with the antibody in the serum by using a labeled anti-immunoglobulin (IgG antibody). Also included are diagnostic reagents for detecting glioma containing 1 or more kinds of the whole or partial glioma antigens thus prepared, and/or an antibody binding with the glioma antigens and/or parts of them, probes for detecting or

CC diagnosis of glioma containing the whole or partial antisense chains of DNAs or RNAs encoding the prepared whole or partial glioma antigens, DNAs (or cDNAs) encoding the glioma antigens (AAU9609-AAU9614) and the protein sequences for the antigens (or peptides derived from them), a DNA hybridisable with the DNA in under stringent conditions and encoding a protein with immune induction activity, a fusion protein or peptide formed by bonding the protein or its partial peptide with a marker protein or peptide tag, an antibody for the protein or its partial peptide, a host cell containing an expression system for expressing the peptide, a non-human animal which has deletion of the gene function encoding the protein or its partial peptide on the chromosome, or has overexpression of the protein or its partial peptide and screening substances promoting or inhibiting immune induction activity by using the protein or its partial peptide, the test substances and T cells, and measuring and evaluating immune induction activity in T cells. The antigen and DNA encoding it, are applicable in diagnosis and treatment of (e.g. by peptide or gene therapy) glioma, human malignant brain tumour and other cancers, and for studying onset of glioma. The present sequence encodes the Glioma antigen KU-GB-5.

CC XX SQ Sequence 3847 BP; 1197 A; 744 C; 819 G; 1086 T; 0 U; 1 Other;

Query Match Score 20.4; DB 6; Length 3847;  
 Best Local Similarity 95.5%; Pred. No. 7.4;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCCGTTGTACTCC 22  
 Db 30 GTGGAAGCCGCCGTTGTACTCC 9

RESULT 14 AAK51608/c ID AAK51608 standard; cDNA; 4450 BP.  
 XX AC AAK51608;  
 XX DT 06-NOV-2001 (first entry)  
 XX DE Human polynucleotide SEQ ID NO 153.  
 XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haemopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorder; arthritis; inflammation; SS.  
 XX OS Homo sapiens.  
 XX PN WO2001517190-A2.  
 XX PD 09-AUG-2001.  
 XX PF 05-FEB-2001; 2001WO-US004098.  
 XX PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620255.  
 PR 01-SEP-2000; 2000US-00654336.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-0069325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX PA (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y; Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW; Xue AJ, Yang Y, Weijhrman T, Goodrich R;  
 XX DR WPI; 2001-476283/51.  
 DR P-PSDB; AAM78475.  
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful

PT in diagnosis and gene therapy.  
 XX Claim 1; Page 862-867; 6221pp; English.  
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM7823-AAM8032) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibitin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK5282) and 3666 (AAM80020) are omitted from the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX Sequence 4450 BP; 1252 A; 987 C; 1047 G; 1164 T; 0 U; 0 Other;  
 SQ Query Match 92.7%; Score 20.4; DB 4; Length 4450;  
 Best Local Similarity 95.5%; Pred. No. 7.5;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Db 609 GTCGAAGCCGCCGTTGACTCC 22  
 Qy 1 GTCGAAGCCGCCGTTGACTCC 22  
 AC ABA09174 standard; cDNA; 4470 BP.  
 XX RE RESULT 15  
 XX ABA09174/C  
 ID ABA09174  
 DT 11-JAN-2002 (First entry)  
 XX Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.  
 DB Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.  
 XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haemopoiesis; regulation; tissue growth; immunomodulator; activin;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytostatic; osteopathic; vasotropics; cardiotonic; antibacterial;  
 KW antifungal; pulmonary; antilulcer; ss.  
 OS Homo sapiens.  
 XX WO200157188-A2.  
 PN PD 09-AUG-2001.  
 XX PP 05-FEB-2001; 2001WO-US003880.  
 XX PR 03-FEB-2000; 2000US-0049614.  
 PR 27-APR-2000; 2000US-00560815.  
 XX PA (HYSEQ INC.  
 XX PA Tang YT, Liu C, Dirmancan RT;  
 XX DR WPI; 2001-457740/49.  
 DR P-PSDB; ABB11930.  
 XX Human Proteins and DNA encoding sequences useful for preventing, treating  
 PT or ameliorating a medical condition in a mammalian subject e.g. arthritis

PT and cancer.  
 XX Claim 1; Page 813-814; 1963pp; English.  
 XX Sequences ABB10981-ABB12310 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haemostatic, thrombotic or inhibin-related activities;  
 CC immunomodulatory activity; tissue growth activity;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders,  
 CC polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a cDNA encoding a  
 CC novel human polypeptide of the invention  
 XX Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;  
 SQ Query Match 92.7%; Score 20.4; DB 4; Length 4470;  
 Best Local Similarity 95.5%; Pred. No. 7.5;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GTCGAAGCCGCCGTTGACTCC 22  
 Db 609 GTCGAAGCCGCCGTTGACTCC 588

Search completed: July 18, 2004, 12:45:05  
 Job time : 37.7148 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:32:20 ; Search time 7.42951 Seconds  
(without alignments)  
1643.301 Million cell updates/sec

Title: US-10-051-769-6  
Perfect score: 22  
Sequence: 1 gtggaggccgcgcgttgtactcc 22

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:  
 1: /cnm2\_6/podata/2/ina/5A COMB.seq/\*  
 2: /cnm2\_6/podata/2/ina/5B COMB.seq/\*  
 3: /cnm2\_6/podata/2/ina/6A COMB.seq/\*  
 4: /cnm2\_6/podata/2/ina/6B COMB.seq/\*  
 5: /cnm2\_6/podata/2/ina/PCTUS COMB.seq/\*  
 6: /cnm2\_6/podata/2/ina/backtFiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.4	92.7	4303	4	US-09-976-194-809 Sequence 899, App
C 2	18.8	85.5	960	2	US-09-976-194-809 Sequence 1, Appli
C 3	18.8	85.5	960	5	FCT-US93-01034-1 Sequence 1, Appli
C 4	17.2	78.2	681	2	US-08-347-194A-3 Sequence 3, Appli
C 5	17.2	78.2	681	3	US-08-663-682-18 Sequence 18, Appli
C 6	17.2	78.2	750	2	US-08-797-689-3 Sequence 3, Appli
C 7	17.2	78.2	750	4	US-09-984-186-3 Sequence 3, Appli
C 8	17.2	78.2	1818	4	US-09-252-991A-2266 Sequence 2266, Ap
C 9	17.2	78.2	3051	4	US-09-252-991A-2343 Sequence 2343, Ap
C 10	17.2	78.2	3156	4	US-09-252-991A-2574 Sequence 2574, Ap
C 11	17.2	78.2	3234	1	US-08-264-534-21 Sequence 31, Appli
C 12	17.2	78.2	3234	1	US-08-083-590A-10 Sequence 10, Appli
C 13	17.2	78.2	3234	1	US-08-465-682-2 Sequence 31, Appli
C 14	17.2	78.2	3234	2	US-08-346-128-31 Sequence 10, Appli
C 15	17.2	78.2	3234	3	US-08-532-384-10 Sequence 31, Appli
C 16	17.2	78.2	3234	3	US-08-893-328-31 Sequence 1, Appli
C 17	17.2	78.2	6153	1	US-08-347-594A-1 Sequence 1, Appli
C 18	17.2	78.2	6153	3	US-08-463-682-2 Sequence 2, Appli
C 19	17.2	78.2	8575	4	US-09-381-261A-2 Sequence 1, Appli
C 20	17.2	78.2	8802	3	US-08-896-449A-1 Sequence 1, Appli
C 21	17.2	78.2	8802	3	US-09-132-552-1 Sequence 1, Appli
C 22	16.4	74.5	2064	1	US-08-343-128-1 Sequence 1, Appli
C 23	16.2	73.6	237	4	US-09-328-352-888 Sequence 5335, App
C 24	16.2	73.6	441	4	US-09-252-991A-5335 Sequence 5335, App
C 25	16.2	73.6	486	4	US-09-252-991A-13723 Sequence 13723, A
C 26	16.2	73.6	693	4	US-09-252-991A-3399 Sequence 3399, Ap
C 27	16.2	73.6	798	4	US-09-252-991A-3359 Sequence 3359, Ap

**ALIGNMENTS**

```

RESULT 1
US-09-976-594-899/c
; Sequence 899, Application US/09976594
; Patent No. 6677349
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; INVENTOR: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PBRL Program
; SEQ ID NO: 899
; LENGTH: 4303
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: misc feature
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1
US-09-976-594-899

```

```

Query Match 92.7%; Score 20.4; DB 4; Length 4303;
Best Local Similarity 95.5%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTGGAAGCCGCCGTTGACTCC 22
Db 322 GTGGAAGCCGCCGTTGACTCC 301

```

**RESULT 2**  
US-07-841-591A-1/c

```

; Sequence 1, Application US/07841591A
; Patent No. 5900416
; GENERAL INFORMATION:
; APPLICANT: Rugeri, Zaverio M. and Ware, Jerry inventors
; APPLICANT: Ware, Jerry inventors on behalf of The Scripps Research Institute
; TITLE OF INVENTION: Therapeutic Domains of von Willebrand Factor
; TITLE OF INVENTION: von Willebrand Factor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute
; STREET: 10666 No. 5900476th Torrey Pines Road
; CITY: La Jolla
; STATE: California

```

COUNTRY: United States  
 ZIP: 92037  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 1.2 megabyte 5 1/4" Floppy  
 COMPUTER: AST Bravo 386SX  
 OPERATING SYSTEM: MS DOS version 3.2  
 SOFTWARE: WordPerfect 5.1 conv. to ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/841,591A  
 FILING DATE: 26-Feb-92  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA: This is a c-i-p of  
 APPLICATION NUMBER: PCT/US91/07756  
 FILING DATE: 17-Oct-91  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Barron, Alexis  
 REFERENCE/DOCKET NUMBER: 22,702  
 TELECOMMUNICATION INFORMATION:  
 TELEFAX: (215) 923-4466  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 960  
 TYPE: Nucleic Acid  
 STRANDEDNESS: single stranded  
 TOPOLOGY: Linear  
 US-07-841-591A-1

Query Match 85.5%; Score 18.8; DB 2; Length 960;  
 Best Local Similarity 90.9%; Pred. No. 7.8;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTGGAAGGCCGCTGTACTCC 22  
 Db 399 GTGGAAGGCCGCTGTACTCC 378

RESULT 3  
 PCT-US93-02034-1/C  
 Sequence 1, Application PC/TU9302034  
 GENERAL INFORMATION:  
 APPLICANT: Ruggeri, Zaverio M. and  
 APPLICANT: Ware, Jerry, Inventors  
 APPLICANT: on behalf of The Scripps Research  
 APPLICANT: Institute  
 TITLE OF INVENTION: Therapeutic Domains of  
 TITLE OF INVENTION: von Willebrand Factor  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: The Scripps Research Institute  
 STREET: 10666 North Torrey Pines Road  
 CITY: La Jolla  
 STATE: California  
 COUNTRY: United States  
 ZIP: 92037

COMPUTER READABLE FORM:  
 MEDIUM TYPE: 2 megabye 3.25" Floppy  
 COMPUTER: AST Bravo 386SX  
 OPERATING SYSTEM: MS DOS version 3.2  
 SOFTWARE: WordPerfect 5.1 conv. to ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/02034  
 FILING DATE: 11930223  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA: This is a c-i-p of  
 APPLICATION NUMBER: PCT/US91/07756  
 FILING DATE: 17-Oct-91  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Barron, Alexis  
 REFERENCE/DOCKET NUMBER: 22,702  
 TELECOMMUNICATION INFORMATION:

Query Match 85.5%; Score 18.8; DB 2; Length 960;  
 Best Local Similarity 90.9%; Pred. No. 7.8;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTGGAAGGCCGCTGTACTCC 22  
 Db 399 GTGGAAGGCCGCTGTACTCC 378

RESULT 4  
 US-08-347-594A-3/C  
 Sequence 3, Application US/08347594A  
 Patent No. 5849536  
 GENERAL INFORMATION:  
 APPLICANT: Garfinkel, Leonard  
 APPLICANT: Richter, Tamara  
 TITLE OF INVENTION: CLONING AND PRODUCTION OF HUMAN VON  
 TITLE OF INVENTION: WILLBRAND FACTOR GP1b BINDING DOMAIN POLYPEPTIDES AND  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: John P. White  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/347,594 A  
 FILING DATE: No. 5849536ember 30, 1994  
 CLASSIFICATION: 425  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P.  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 36337-B2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-278-0400  
 TELEX: 212-391-0525  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 681 base Pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..681  
 US-08-347-594A-3

Query Match 78.2%; Score 17.2; DB 2; Length 681;  
 Best Local Similarity 86.4%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGGAAGGCCGCTGTACTCC 22  
 Db 183 GTGGAAGGCCGCTGTACTCC 162

RESULT 5  
US-08-463-682-18/c  
Sequence 18, Application US/08463682  
Patent No. 6008193  
GENERAL INFORMATION:  
APPLICANT: Leonard Garfinkel, et al.  
TITLE OF INVENTION: Cloning and Production of Human Von Willebrand Factor GP1b Binding Domain Polypeptides and Methods of Using Same  
TITLE OF INVENTION: Willebrand Factor GP1b Binding Domain Polypeptides and Methods of Using Same  
NUMBER OF SEQUENCES: 24  
APPLICATION NUMBER: US/08/463,682  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 36537-B2-Y  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 651 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: nucleic acid  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..681  
US-08-463-682-18

Query Match 78.2%; Score 17.2; DB 3; Length 681;  
Best Local Similarity 86.4%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 6  
US-08-737-689-3/c  
Sequence 3, Application US/08737689  
Patent No. 5876969  
GENERAL INFORMATION:  
APPLICANT: Flier, Reinhard  
APPLICANT: Fournier, Alain  
APPLICANT: Guittot, Jean-Dominique  
APPLICANT: Jung, Gerard  
APPLICANT: Yeh, Patrice  
TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES, COMPOSITION PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION CONTAINING SAID POLYPEPTIDES  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Road,  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

STREET: 500 Arcola Road, 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426  
GENERAL INFORMATION:  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: System 7.1  
SOFTWARE: Word 5.1 (Patentin)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/797,689  
FILING DATE: 31-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/256,927  
FILING DATE: 28-JUL-1994  
APPLICATION NUMBER: FR 92/01064  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00085  
FILING DATE: 28-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Smich Ph.D., Julie K.  
REGISTRATION NUMBER: P-38,619  
REFERENCE/DOCKET NUMBER: ST92006-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 454-3839  
TELEFAX: (610) 454-3808  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 750 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..746  
US-08-797-689-3

Query Match 78.2%; Score 17.2; DB 2; Length 750;  
Best Local Similarity 86.4%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCCGCTGTACTCC 22  
Db 183 GTGGAGCCGCTGGTACTCC 162

Query Match 78.2%; Score 17.2; DB 3; Length 681;  
Best Local Similarity 86.4%; Pred. No. 42;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 7  
US-09-984-186-3/c  
Sequence 3, Application US/09984186  
Patent No. 6686179  
GENERAL INFORMATION:  
APPLICANT: Flier, Reinhard  
APPLICANT: Fournier, Alain  
APPLICANT: Guittot, Jean-Dominique  
APPLICANT: Jung, Gerard  
APPLICANT: Yeh, Patrice  
TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES, COMPOSITION PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION CONTAINING SAID POLYPEPTIDES  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Road,  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

RESULT 8  
US-08-737-689-3/c  
Sequence 3, Application US/08737689  
Patent No. 5876969  
GENERAL INFORMATION:  
APPLICANT: Flier, Reinhard  
APPLICANT: Fournier, Alain  
APPLICANT: Guittot, Jean-Dominique  
APPLICANT: Jung, Gerard  
APPLICANT: Yeh, Patrice  
TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES, COMPOSITION PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION CONTAINING SAID POLYPEPTIDES  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.

COMPUTER: Macintosh  
 OPERATING SYSTEM: System 7.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/984,186  
 FILING DATE: 29-OCT-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/797,689  
 FILING DATE: 31-JAN-1997  
 APPLICATION NUMBER: US/08/256,927  
 FILING DATE: 28-JUL-1994  
 APPLICATION NUMBER: FR 92/01064  
 FILING DATE: 31-JAN-1992  
 APPLICATION NUMBER: PCT/FR93/00095  
 FILING DATE: 28-JAN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Smith Ph. D., Julie K.  
 REGISTRATION NUMBER: P-38,619  
 REFERENCE/DOCKET NUMBER: ST20006-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (610) 454-3339  
 TELEFAX: (610) 454-3808  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 750 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 3..746  
 SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
 US-09-984-186-3

Query Match Score 78.2%; DB 4; Length 750;  
 Best Local Similarity 86.4%; Pred. No. 42;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0; Length 10  
 US-09-252-991A-2574/C

RESULT 9  
 US-09-252-991A-2343  
 ; Sequence 2343, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIORITY NUMBER: US 60/074,788  
 ; PRIORITY FILING DATE: 1998-02-18  
 ; PRIORITY APPLICATION NUMBER: US 60/094,190  
 ; PRIORITY FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO: 2343  
 ; LENGTH: 3051  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (2880)  
 ; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
 US-09-252-991A-2343

Query Match Score 78.2%; DB 4; Length 3051;  
 Best Local Similarity 86.4%; Pred. No. 44;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0; Length 3156  
 US-09-252-991A-2574

Query Match Score 78.2%; DB 4; Length 3051;  
 Best Local Similarity 86.4%; Pred. No. 44;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0; Length 3156  
 US-09-252-991A-2574

RESULT 8  
 US-09-252-991A-2266  
 ; Sequence 2266, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIORITY NUMBER: US 60/074,788  
 ; PRIORITY FILING DATE: 1998-02-18  
 ; PRIORITY APPLICATION NUMBER: US 60/094,190  
 ; PRIORITY FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO: 2266  
 ; LENGTH: 3156  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1027)  
 ; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
 US-09-252-991A-2266

Query Match Score 78.2%; DB 4; Length 1818;  
 Best Local Similarity 86.4%; Pred. No. 43;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0; Length 1818  
 US-09-252-991A-2266

Query Match Score 78.2%; DB 4; Length 1951;  
 Best Local Similarity 86.4%; Pred. No. 43;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0; Length 1951  
 US-09-252-991A-2266

RESULT 11  
 US -08-264-534-31  
 Sequence 10-A Application US/08264534  
 Patent No. 5648464  
 GENERAL INFORMATION:  
 i APPLICANT: Artavanis-Tsalakonas, Spyridon et al.  
 i TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains  
 i TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon  
 i NUMBER OF SEQUENCES: 34  
 i CORRESPONDENCE ADDRESS:  
 i ADDRESSEE: Pennie & Edmonds  
 i STREET: 1155 Avenue of the Americas  
 i CITY: New York  
 i STATE: New York  
 i COUNTRY: U.S.A.  
 i ZIP: 10036  
 COMPUTER READABLE FORM:  
 i MEDIUM TYPE: Floppy disk  
 i COMPUTER: IBM PC compatible  
 i OPERATING SYSTEM: PC-DOS/MS-DOS  
 i SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 i APPLICATION NUMBER: US/08/083,590A  
 i FILING DATE: 25-JUN-1993  
 i CLASSIFICATION: 435  
 i ATTORNEY/AGENT INFORMATION:  
 i NAME: Misrock, S. Leslie  
 i REGISTRATION NUMBER: 18,872  
 i REFERENCE/DOCKET NUMBER: 7326-015  
 i TELECOMMUNICATION INFORMATION:  
 i TELEPHONE: 212 790-9090  
 i TELEFAX: 212 8698864/9741  
 i TELEX: 66141 PENNIE  
 i INFORMATION FOR SEQ ID NO: 10:  
 i SEQUENCE CHARACTERISTICS:  
 i LENGTH: 3234 base pairs  
 i TYPE: nucleic acid  
 i STRANDEDNESS: double  
 i TOPOLOGY: unknown  
 i MOLECULE TYPE: cDNA  
 i FEATURE:  
 i NAME/KEY: CDS  
 i LOCATION: 1..3234  
 US-08-083-590A-10

Query Match  
 Best Local Similarity 78.2%; Score 17.2%; DB 1; Length 3234;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 13  
 US-08-465-500-31  
 Sequence 31, Application US/08465500  
 Patent No. 5789195  
 GENERAL INFORMATION:  
 i APPLICANT: Artavanis-Tsalakonas, Spyridon  
 i APPLICANT: Muskavitch, Marc A.T.  
 i APPLICANT: Fehon, Richard G.  
 i APPLICANT: Reby, Clari  
 i APPLICANT: Blaumuller, Cristine M.  
 i APPLICANT: Shepard, Scott B.  
 i TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS  
 i TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON  
 i NUMBER OF SEQUENCES: 34  
 i CORRESPONDENCE ADDRESS:  
 i STREET: 1155 Avenue of the Americas  
 i CITY: New York  
 i STATE: NY  
 i COUNTRY: USA  
 i ZIP: 10036-2711

COMPUTER READABLE FORM:  
 i MEDIUM TYPE: Floppy disk  
 i COMPUTER: IBM PC compatible  
 i OPERATING SYSTEM: PC-DOS/MS-DOS  
 i SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 i APPLICATION NUMBER: US/08/465,500  
 i FILING DATE: 05-JUN-1995  
 i CLASSIFICATION: 435  
 i ATTORNEY/AGENT INFORMATION:  
 i NAME: Misrock, S. Leslie  
 i REGISTRATION NUMBER: 18,872  
 i REFERENCE/DOCKET NUMBER: 7326-034

RESULT 12  
 US-08-083-590A-10  
 Sequence 10-A Application US/08083590A  
 Patent No. 5786158  
 GENERAL INFORMATION:  
 i APPLICANT: Artavanis-Tsalakonas, S. et al.  
 i TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
 i TITLE OF INVENTION: And Compositions Based On No. 5786158th Proteins And  
 i TITLE OF INVENTION: Nucleic Acids  
 i NUMBER OF SEQUENCES: 21  
 i CORRESPONDENCE ADDRESS:  
 i ADDRESSEE: Pennie & Edmonds  
 i STREET: 1155 Avenue of the Americas  
 i CITY: New York  
 i STATE: New York  
 i COUNTRY: U.S.A.

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9010  
 TELEFAX: (212) 869-8864/9741  
 INFORMATION FOR SEQ ID NO: 31:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3234 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: unknown  
 MOLECULE TYPE: cDNA  
 US-08-465-500-31

Query Match 78.2%; Score 17.2; DB 1; Length 3234;  
 Best Local Similarity 86.4%; Pred. No. 44;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCCGTGTACTCC 22  
 Db 1762 GTGGATGCCGAGTGTGCTCC 1783

RESULT 14  
 US-08-346-128-31  
 Sequence 31; Application US/08346128  
 Patent No. 5856441

GENERAL INFORMATION:  
 APPLICANT: Artavanis-Tsalikas, Spyridon et al.  
 TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains  
 TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon  
 NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent.In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/346,128  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIORITY APPLICATION NUMBER:  
 APPLICATION NUMBER: 08/083,590  
 FILING DATE: 25-JUN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Misrock, S. Leslie  
 ZIP: 10036  
 REGISTRATION NUMBER: 18,872  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212 790-9090  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3234 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: unknown  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..3234  
 US-08-532-384-10

Query Match 78.2%; Score 17.2; DB 3; Length 3234;  
 Best Local Similarity 86.4%; Pred. No. 44;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCCGTGTACTCC 22  
 Db 1762 GTGGATGCCGAGTGTGCTCC 1783

Search completed: JULY 18, 2004, 14:38:30  
 Job time : 8.42951 secs

Query Match 78.2%; Score 17.2; DB 2; Length 3234;  
 Best Local Similarity 86.4%; Pred. No. 44;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCCGTGTACTCC 22  
 Db 1762 GTGGATGCCGAGTGTGCTCC 1783

Query Match 78.2%; Score 17.2; DB 2; Length 3234;  
 Best Local Similarity 86.4%; Pred. No. 44;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCCGTGTACTCC 22  
 Db 1762 GTGGATGCCGAGTGTGCTCC 1783

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:36:30 ; Search time 38 5902 Seconds  
 (without alignments)  
 2781.271 Million cell updates/sec

Title: US-10-051-769-6

Perfect score: 22

Sequence: 1 gtggaaaggcccggtgtactcc 22

Scoring table: IDENTITY\_NUC  
 Gapop 10\_0 , Gapext 1.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
 Listing first 45 summaries

Published Applications NA.\*

Database : 1: /cgn2\_6/ptodata/2/pubpnas/US07\_PUBCOMB.seq:  
 2: /cgn2\_6/ptodata/2/pubpnas/US07\_PUBCOMB.seq:  
 3: /cgn2\_6/ptodata/2/pubpnas/US07\_PUBCOMB.seq:  
 4: /cgn2\_6/ptodata/2/pubpnas/US07\_PUBCOMB.seq:  
 5: /cgn2\_6/ptodata/2/pubpnas/US07\_PUBCOMB.seq:  
 6: /cgn2\_6/ptodata/2/pubpnas/US07\_PUBCOMB.seq:  
 7: /cgn2\_6/ptodata/2/pubpnas/US08\_NEW\_PUB.seq:  
 8: /cgn2\_6/ptodata/2/pubpnas/US08\_PUBCOMB.seq:  
 9: /cgn2\_6/ptodata/2/pubpnas/US08\_PUBCOMB.seq:  
 10: /cgn2\_6/ptodata/2/pubpnas/US09\_PUBCOMB.seq:  
 11: /cgn2\_6/ptodata/2/pubpnas/US09C\_PUBCOMB.seq:  
 12: /cgn2\_6/ptodata/2/pubpnas/US09\_NEW\_PUB.seq:  
 13: /cgn2\_6/ptodata/2/pubpnas/US09\_NEW\_PUB.seq:  
 14: /cgn2\_6/ptodata/2/pubpnas/US10A\_PUBCOMB.seq:  
 15: /cgn2\_6/ptodata/2/pubpnas/US10B\_PUBCOMB.seq:  
 16: /cgn2\_6/ptodata/2/pubpnas/US10C\_PUBCOMB.seq:  
 17: /cgn2\_6/ptodata/2/pubpnas/US60\_NEW\_PUB.seq:  
 18: /cgn2\_6/ptodata/2/pubpnas/US60\_NEW\_PUB.seq:  
 19: /cgn2\_6/ptodata/2/pubpnas/US60\_PUBCOMB.seq:  
 Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Match Length	DB ID	Description
1	22	100.0	22	15 US-10-051-769-6
2	22	100.0	22	15 US-10-224-624-6
c 3	22	100.0	263	15 US-10-051-769-4
c 4	20.4	92.7	261	15 US-10-051-769-2
c 5	20.4	92.7	261	15 US-10-224-624-4
c 6	20.4	92.7	261	15 US-10-224-624-2
c 7	20.4	92.7	873	15 US-10-051-769-3
c 8	20.4	92.7	873	15 US-10-224-624-3
c 9	20.4	92.7	960	15 US-10-051-769-1
c 10	20.4	92.7	960	15 US-10-224-624-1
c 11	20.4	92.7	2639	16 US-10-104-047-684
c 12	20.4	92.7	3465	16 US-10-224-624-9
c 13	20.4	92.7	3649	13 US-10-112-944-63
c 14	20.4	92.7	3832	15 US-10-224-624-7

**ALIGNMENTS**

RESULT 1  
 US-10-051-769-6

; Sequence 6, Application US/10051769  
 ; Publication No. US2003044811A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCKINNON, Randy D.  
 ; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
 ; CURRENT APPLICATION NUMBER: US/10/051,769  
 ; FILE REFERENCE: 268/260 (RWJ-00-37)  
 ; CURRENT FILING DATE: 2001-10-20  
 ; PRIOR APPLICATION NUMBER: US 60/242,160  
 ; PRIOR FILING DATE: 2000-10-20  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 6  
 ; LENGTH: 22;  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic Primer

US-10-051-769-6



RESULT 7  
US-10-051-769-3/c  
Sequence 3, Application US/10051769  
Publication No. US20030044811A1  
GENERAL INFORMATION:  
APPLICANT: MCKINNON, Randy D.  
TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
FILE REFERENCE: 268/260 (RWJ-00-37)  
CURRENT APPLICATION NUMBER: US/10/051,769  
CURRENT FILING DATE: 2001-10-20  
PRIOR APPLICATION NUMBER: US 60/242,160  
PRIOR FILING DATE: 2000-10-20  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 873  
TYPE: DNA  
ORGANISM: Homo Sapiens

US-10-051-769-3  
Query Match 92.7%; Score 20.4; DB 15; Length 873;  
Best Local Similarity 95.5%; Pred. No. 1.8;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GTGAAAGCCCGTGTACTCC 22  
Db 639 GTGAAAGCCCCGGTGTACTCC 588

RESULT 8  
US-10-224-624-3/c  
Sequence 3, Application US/10224624  
Publication No. US20030108915A1  
GENERAL INFORMATION:  
APPLICANT: MCKINNON, Randall D.  
TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN  
FILE REFERENCE: 54704, 8059, US00  
CURRENT APPLICATION NUMBER: US/10/224,624  
CURRENT FILING DATE: 2002-08-10  
PRIOR APPLICATION NUMBER: 60/242,160  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 10/051,769  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 960  
TYPE: DNA  
ORGANISM: Homo Sapiens

US-10-224-624-1  
Query Match 92.7%; Score 20.4; DB 15; Length 960;  
Best Local Similarity 95.5%; Pred. No. 1.8;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GTGAAAGCCCGTGTACTCC 22  
Db 639 GTGAAAGCCCCGGTGTACTCC 618

RESULT 11  
US-10-104-047-684/C  
Sequence 684, Application US/10104047  
Publication No. US20030236392A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: NO. US20030236392A1el full length cDNA  
FILE REFERENCE: HL-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 684  
LENGTH: 2639  
TYPE: DNA  
ORGANISM: Homo sapiens

US-10-104-047-684  
Query Match 92.7%; Score 20.4; DB 16; Length 2639;  
Best Local Similarity 95.5%; Pred. No. 1.8;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGAAAGCCCGTGTACTCC 22

Db 822 GTCGAAGCCCGCTTGACTCC 801

RESULT 12  
US-10-224-624-9/c  
Sequence 9, Application US/10224624  
Publication No. US20030108915A1  
GENERAL INFORMATION:  
APPLICANT: MCKINNON, Randall D.  
TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLI1EN  
CURRENT APPLICATION NUMBER: US/10/224,624  
FILE REFERENCE: 54104.8059.US00  
PRIORITY FILING DATE: 2002-08-20  
PRIOR APPLICATION NUMBER: 60/242,160  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 10/051,769  
PRIOR FILING DATE: 2001-10-20  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 9  
LENGTH: 3465  
TYPE: DNA  
ORGANISM: Homo sapiens

US-10-224-624-9

Query Match 92.7%; Score 20.4; DB 15; Length 3465;  
Best Local Similarity 95.5%; Pred. No. 1.8;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGAAGCCCGCTTGACTCC 22  
Db 609 GTCGAAGCCCGCTTGACTCC 588

---

RESULT 13  
US-10-112-944-63/c  
Sequence 63, Application US/10112944  
Publication No. US20040048249A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Yang, Yonghong  
APPLICANT: Weng, Gezhi  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Xue, Aidong J.  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Wehrman, Tom  
APPLICANT: Ghosh, Malabika  
APPLICANT: Wang, Dunrui  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wang, Zhiwei  
TITLE OF INVENTION: Novel Nucleic Acids and  
Title of Invention: Secreted Polypeptides  
FILE REFERENCE: 805A  
CURRENT APPLICATION NUMBER: US/10/112,944  
CURRENT FILING DATE: 2002-03-28  
PRIOR APPLICATION NUMBER: US 09/488,725  
PRIOR FILING DATE: 2000-01-21  
PRIOR APPLICATION NUMBER: US 09/491,404  
PRIOR FILING DATE: 2000-01-25  
PRIOR APPLICATION NUMBER: US 09/496,914  
PRIOR FILING DATE: 2000-02-03  
PRIOR APPLICATION NUMBER: US 09/515,126  
PRIOR FILING DATE: 2000-02-28  
PRIOR APPLICATION NUMBER: US 09/519,705  
PRIOR APPLICATION NUMBER: US 09/540,217  
PRIOR FILING DATE: 2000-03-31  
PRIOR APPLICATION NUMBER: US 09/552,929  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: US 09/577,408  
PRIOR FILING DATE: 2000-05-18

---

RESULT 14  
US-10-224-624-7/c  
Sequence 7, Application US/10224624  
Publication No. US20030108915A1  
GENERAL INFORMATION:  
APPLICANT: MCKINNON, Randall D.  
TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLI1EN  
FILE REFERENCE: 54104.8059.US00  
CURRENT APPLICATION NUMBER: US/10/224,624  
PRIORITY FILING DATE: 2002-08-20  
PRIOR APPLICATION NUMBER: 60/242,160  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 10/051,769  
PRIOR FILING DATE: 2001-10-20  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 7  
LENGTH: 3832  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1) .. (3462)  
US-10-112-944-63

Query Match 92.7%; Score 20.4; DB 15; Length 3832;  
Best Local Similarity 95.5%; Pred. No. 1.8;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGAAGCCCGCTTGACTCC 22  
Db 786 GTCGAAGCCCGCTTGACTCC 765

---

RESULT 15  
US-10-250-823-11/c  
Sequence 11, Application US/10250823  
Publication No. US20040072265A1  
GENERAL INFORMATION:  
APPLICANT: KEIO UNIVERSITY  
TITLE OF INVENTION: Human Brain Tumor Antigens and method for preparation  
FILE REFERENCE: P0000114  
CURRENT APPLICATION NUMBER: US/10/250,823  
CURRENT FILING DATE: 2003-07-07  
PRIOR APPLICATION NUMBER: JP 2001-01-09  
PRIOR FILING DATE: 2001-01-09  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 11  
LENGTH: 3847  
TYPE: DNA  
ORGANISM: Homo sapiens

```
? FEATURE:  
?   NAME/KEY: CDS  
?   LOCATION: (211)..(2833)  
US-10-250-823-11  
  
Query Match 92.7%; Score 20.4; DB 12; Length 3847;  
Best Local Similarity 95.5%; Pred. No. 1.8;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY      1 GTGGAAGCCGGTGTACTCC 22  
       ||||| ||| ||| ||| ||| |||  
Db      30 GTGGAACCCCGGTGTACTCC 9
```

Search completed: July 18, 2004, 14:47:32  
Job time : 39.5902 secs



GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:55:38 ; Search time 272.872 Seconds  
 (without alignments)  
 2407.604 Million cell updates/sec

Title: US-10-051-769-6  
 Perfect score: 22  
 Sequence: 1 gtcgaaggccgcgttgtactcc 22

Scoring table: IDENTITY\_NUC  
 Gapop 1.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:  
 Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : EST:  
 1: em\_estba:  
 2: em\_estbm:  
 3: em\_estin:  
 4: em\_estml:  
 5: em\_estov:  
 6: em\_estpl:  
 7: em\_estro:  
 8: em\_htc:  
 9: gb\_est1:  
 10: gb\_est2:  
 11: gb\_htc:  
 12: gb\_est3:  
 13: gb\_est4:  
 14: gb\_est5:  
 15: em\_estm:  
 16: em\_eston:  
 17: em\_gss\_hum:  
 18: em\_gss\_inv:  
 19: em\_gss\_pln:  
 20: em\_gss\_vrt:  
 21: em\_gss\_fun:  
 22: em\_gss\_mam:  
 23: em\_gss\_mus:  
 24: em\_gss\_pro:  
 25: em\_gss\_fod:  
 26: em\_gss\_phg:  
 27: em\_gss\_vrl:  
 28: gb\_gss1:  
 29: gb\_gss2:  
 30: em\_gss\_vrl:  
 31: em\_gss\_vrl:  
 32: em\_gss\_vrl:  
 33: em\_gss\_vrl:  
 34: em\_gss\_vrl:  
 35: em\_gss\_vrl:  
 36: em\_gss\_vrl:  
 37: em\_gss\_vrl:  
 38: em\_gss\_vrl:  
 39: em\_gss\_vrl:  
 40: em\_gss\_vrl:  
 41: em\_gss\_vrl:  
 42: em\_gss\_vrl:  
 43: em\_gss\_vrl:  
 44: em\_gss\_vrl:  
 45: em\_gss\_vrl:

## ALIGNMENTS

RESULT 1  
 BU449181/C  
 LOCUS 60376743 OF1 CSEQRBN13 Gallus gallus mRNA clone ChEST69518 5', mRNA  
 DEFINITION sequence.  
 BU449181  
 ACCESSION BU449181.1  
 VERSION GI:25938492  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 Archosuria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 730)  
 AUTHORS Boardman, P.B., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 2233554  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409

Result No.	Score	Query	Match	Length	DB ID	Description
c 1	22	100.0	730	13	BU449181	603767430
c 2	22	100.0	898	13	BU453045	603229082
c 3	22	100.0	1145	28	CC290954	CH261-100
c 4	20.4	92.7	437	13	BY276266	BY276266

**SUMMARIES**

%

FEATURES source Email: Simon.Hubbard@umist.ac.uk.

1. .730 /organism="Gallus gallus"  
   /mol\_type="mRNA"  
   /strain="Layer"  
   /db\_xref="taxon:9031"  
   /clone="ChEBI69518"  
   /sex="Female"  
   /dev\_stage="adult"  
   /lab\_host="DH10B"  
   /clone\_lib="CSBQBN13"  
   /note="Organ: ovary; Vector: pBluescript II KS(+); site\_1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones; cDNA synthesis was initiated using an Oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adaptors, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al. PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996) : 791, except that a significantly longer reannealing hybridization was used."

ORIGIN Query Match 100.0% Score 22; DB 13; Length 898;  
   Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
   Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 100.0% Score 22; DB 13; Length 898;  
   Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
   Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3 CC290954 LOCUS CH261-100B2\_Sp6.1 CH261\_100B2\_0033 DEFINITION Genomic survey sequence. GSS CC290954 VERSION CC290954.1 GI:30662395 SOURCE Gallus gallus (chicken)  
   ORGANISM Gallus gallus  
   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus;  
   1 (bases 1 to 1145)  
   REFERENCES Kremitzer, C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J., Warren, W., Graves,T., Mardis,E. and Wilson,R.  
   TITLE Gallus gallus BAC End Reads  
   JOURNAL Unpublished (2003)  
   COMMENT Contact: Richard K. Wilson  
   Email: submissions@watson.wustl.edu  
   Insert Length: 182000 Std Error: 0.00  
   Seq Primer: Sp6 ATTAGTGCACTATAG  
   Class: BAC ends  
   High quality sequence start: 28  
   High quality sequence stop: 628  
   FEATURES source  
   1..1145  
   /organism="Gallus gallus"  
   /mol\_type="genomic DNA"  
   /strain="Red Jungle Fowl"  
   /db\_xref="taxon:9031"  
   /clone="CH261-100B2"  
   /sex="female"  
   /cell\_line="UCD001, inbred 256"  
   /clone\_lib="CH261"  
   /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CH261 Female Chicken library - For library and clone ordering information: http://www.chori.org/bacpac"

ORIGIN Query Match 100.0% Score 22; DB 28; Length 1145;  
   Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
   Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES source Email: Simon.Hubbard@umist.ac.uk.

1. .898 /organism="Gallus gallus"  
   /mol\_type="mRNA"  
   /strain="Layer"  
   /db\_xref="taxon:9031"  
   /clone="ChEBI210C22"  
   /sex="Female"

Qy	1	GTGGAAGCCCGCGTGTACTTC	22	
Db	527	GTGGAAGCCCGCGTGTACTTC	548	
RESULT	4			
BY276266/C	BY276266	437 bp	mRNA linear	EST 11-DEC-2002
DEFINITION	RIKEN full-length enriched, visual cortex Mus musculus mRNA clone KA3031851 5'			
ACCESSION	BY276266			
VERSION	BY276266..1			
KEYWORDS	EST			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.			
REFERENCE	I (bases 1 to 437)			
AUTHORS	Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamada,I., Tomaru,Y., Hasegawa,Y., Nogami,A., Kiyosawa,H., Yagi,K., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batyalov,S., Beisel,K.W., Blake,J.A., Brusic,V., Bradford,D., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Fraser,K.S., Gaasterland,T., Grommond,S., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Jarvis,E.D., Kansai,A., Gustincic,S., Hirokawa,N., Jackson,I.J., King,B.L., Konagaya,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., Lyons,P.A., Maglott,D.R., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Malrais,L., Marchionni,L., McKenzie,L., Mikici,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Peretea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sutana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlenstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yang,Y., Yang,J., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawaji,U., Azawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Iishi,Y., Ichii,M., Kaewa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E., and Hayashizaki,Y.			
RESULT	5			
BB638573/C	BB638573	458 bp	mRNA linear	EST 31-AUG-2001
LOCUS	RIKEN full-length enriched, 3 days neonate thymus Mus			
DEFINITION	musculus cDNA clone A630041P07 5', mRNA sequence.			
BB638573	BB638573			
BB638573.1	BB638573.1			
EST.				
Qy	1	GTGGRAGCGCGCGTGTACTCC	22	
Db	324	GTGCAGAACGCCGCGTGTACTCC	303	
ORIGIN				
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
		/Location/Qualifiers		
		1..437		
		/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..437	/organism="Mus musculus"		
		/mol_type="mRNA"		
		/strain="C57BL/6J"		
		/db_xref=taxon:10090		
		/clone_id="KA3031851"		
		/tissue_type="visual cortex"		
		/clone_id="RIKEN full-length enriched, visual cortex"		
FEATURES				
source	1..43			

and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
 Sugahara, Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
 Yamamoto, I., Kiysawa, H., Kondo, S., Saito, T., Shinagawa, A.,  
 Aizawa, K., Fukuda, S., Hara, Y., and Hayashizaki, Y.  
 Mapping of 19032 mouse cDNAs on mouse chromosomes. *J. Struct. Funct. Genomics* 2 (Pre), L72-L6 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.  
 Location/Qualifiers  
 1. • 458  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10900"  
 /clone="A63041P07"  
 /tissue\_type="thymus"  
 /dev\_stage="3 days neonate"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, 3 days neonate thymus"  
 /note="Site 1: Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',

TITLE	Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M., and Hayashizaki, Y.	
	JOURNAL	RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. <i>Genome Res.</i> 10 (10), 1617-1630 (2000) wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Togawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline on 384 multicapillary sequencer. <i>Genome Res.</i> 10 (11), 1757-1771 (2000) Konno, H., Fukunishi, Y., Shibata, K., Itch, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. <i>Genome Res.</i> 11 (2), 281-289 (2001) Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamamoto, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. <i>Mamm. Genome</i> 12, 673-677 (2001) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> ) for further details. e mouse tissues.	
FEATURES	source	Location/Qualifiers
		1. .464 /organism="Mus musculus" /moi_type="mRNA" /db_xref="taxon:10090" /clone="D030000C17" /dev_stage="9 days embryo" /lab_host="DH10B" /clone_lib="RIKEN full-length enriched, 9 days embryo" /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGGAGATTCGAGTTAAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XbaI. Vector: a modified pBluescript KS (+) after bulk excision from Lambda FLC I."
ORIGIN	Query Match 92.7%; Score 20.4; DB 10; Length 464; Best Local Similarity 95.5%; Pred. No. 5.4e+02; Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	1 GTGCAAGCGCCGTGTACTCC 22 75 GTGAAACGGCCGTGTACTCC 54



TITLE	Mathewson,C., Wye,N., Masson,A., Brown-John,M., Jones,S., Schein,J., Marra,M., de Jong,P., Keele,J.W. and Kappes,S.M.	FEATURES	High quality sequence stop: 603.
JOURNAL	Bovine BAC End Sequences From Library CHORI-240, PLATES 399 to 478	source	Location/Qualifiers 1. .930
COMMENT	Unpublished (2003) Other GSSs: CH240_441017.T7 Contact: Rob Holt Email: rholt@bcgsc.ca	/organism="Mus musculus" /mol type="mRNA" /strain="FVB/N-3" /db_xref="taxon:10090" /clone="IMAGE:6439792" /tissue_type="tumor, biopsy sample" /dev_stage="5 months" /lab_host="DB10B" /clone_lib="NCI CGAP Mam2" /note="Organelle; mammary; Vector: pCMV-SPORT6; Site 1: Sall1; Site 2: NotI; Cloned unidirectionally; Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"	
Sequencing	The British Columbia Cancer Agency Genome Science Centre 600 W 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6 Tel: 604-877-6085 Fax: 604-877-6216		
	Clones are derived from the bovine BAC library CHORI-240 ( <a href="http://www.chori.org/bacpac/bovine240.htm">http://www.chori.org/bacpac/bovine240.htm</a> ). For BAC library availability, please contact Pieter de Jong ( <a href="mailto:pdejong@mail.cho.org">pdejong@mail.cho.org</a> ). Clones may be purchased from BACPAc Resources ( <a href="http://www.chori.org/bacpac/ordering-information.htm">http://www.chori.org/bacpac/ordering-information.htm</a> ). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Sciences Centre, Canada.		
	Plate: 441 row: O column: 17 Seq primer: SP6 Class: BAC ends.		
FEATURES	1. .738 /organism="Bos taurus" /mol type="genomic DNA" /strain="breed: Hereford" /db_xref="taxon:9913" /clone="CH240_441017" /sex="Male" /cell_type="Blood" /clone_id="CHORI-240" /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: EcoRI-240 Bovine BAC library (Male) produced by Pieter de Jong"		
source		RESULT 11 CNS04CSB/c LOCUS CNS04CSB/c DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone sequence. ACCESSION AL284855 VERSION 1 SOURCE GSS; Genome survey sequence. ORGANISM Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopercyli; Percormorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.	
	Qy 1 GTGGAAAGCCGCGTGTACTCC 22 Db 589 GTGGAAAGCCCCGCGTGTACTCC 568	REFERENCE 1 AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W., and Weissbach,J. TITLE JOURNAL Medline PUBLISHED 2000 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)	
FEATURES	Query Match 92.7% Score 20.4; DB 29; Length 738; Best Local Similarity 95.5%; Pred. No. 6.3e+02; Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
source		RESULT 10 BQ960960/c LOCUS BQ960960 DEFINITION AGENCOURT 8955743 NCI CGAP Mam2 Mus musculus cDNA clone IMAGE:6439792 5', mRNA sequence. ACCESSION BQ960960 VERSION 1 SOURCE NIH-MGC http://mgc.ncbi.nlm.nih.gov/ ORGANISM Mus musculus (house mouse) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 930) EST. Mus musculus Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabb-r@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGIC clone distribution information can be found through the I.M.A.G.E. Consortium/LILN at: <a href="http://image1.liln.gov">http://image1.liln.gov</a>	
	Query Match 92.7% Score 20.4; DB 13; Length 930; Best Local Similarity 95.5%; Pred. No. 6.6e+02; Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
FEATURES	1. .958 /organism="Tetraodon nigroviridis"		
source		COMMENT Submitted (12-APR-2000) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY Cedex - FRANCE (E-mail : seqref@genoscope.cnrs.fr) This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.Genoscope.cnrs.fr/Tetraodon">http://www.Genoscope.cnrs.fr/Tetraodon</a> .	

/mol type="genomic DNA"										
/db_xref="taxon:99883"										
/clone="101121"										
/note="Genoscope sequence ID : COBG101AE11LP1-end : T7"										
<b>ORIGIN</b>										
Query Match 92.7%; Score 20.4; DB 29; Length 958;										
Best Local Similarity 95.5%; Pred. No. 6.6e+02;										
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;										
Qy 1 GTGGAGGCCGCGTGTACCC 22										
Db 106 GTGGAGGCCGCGTGTACCC 85										
<b>RESULT 12</b>										
BX341552 LOCUS BX341552 Homo sapiens HEA CELLS COT 25-NORMALIZED Homo sapiens EST.										
ACCESSION BX341552 VERSION BX341552.1										
KEYWORDS SOURCE ORGANISM Homo sapiens										
MATCHES 1 (bases 1 to 989)										
REFERENCE AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.										
TITLE FULL-LENGTH cDNA LIBRARIES AND NORMALIZATION										
JOURNAL Unpublished (2001)										
COMMENT Contact: Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France										
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr										
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5483.r. For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODK002BA06Q1&cluster=5483.r. Contact : Peng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/invitro Corporation 1600 Faraday Avenue Genoscope sequence ID : CSODK002BA06Q1.										
<b>FEATURES source</b>										
1. .989 /organism="Homo sapiens"										
/mol_type="mRNA"										
/db_xref="taxon:96016"										
/clone="CSODK002BA12"										
/cell_type="HELA"										
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"										
/note="1st strand cDNA was primed with a NotI-Oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."										
Query Match 92.7%; Score 20.4; DB 13; Length 989;										
Best Local Similarity 95.5%; Pred. No. 6.7e+02;										
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;										
Qy 1 GTGGAGGCCGCGTGTACCC 22										
Db 876 GTGGAGGCCGCGTGTACCC 855										
<b>RESULT 13</b>										
AK041847/LOCUS										
AK041847 LOCUS										
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630041P07 product:hypothetical PDZ domain (also										

known as DHR or GLGF) containing protein, full insert sequence.

ACCESSION AK041847  
VERSION AK041847.1 GI:26088633  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mammalia; Butheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 Carninci, P. and Hayashizaki, Y.  
AUTHORS Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349656

REFERENCE 2 Carninci, P., Itoh,M., Aizawa,K., Nagaoaka,S., Sasakai,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,T., Fujiiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanuki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--84-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530943  
PUBMED 11076861

REFERENCE 3 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasakai,N., Carninci,P., Konno,H., Furuno,M., Haga,K., Hashizume,W., Hayashizaki,Y., Hayashida,K., Hiramoto,T., Hiraoka,T., Kasukawa,T., Horii,F., Imotani,K., Ishii,Y., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Katoch,H., Kawai,J., Kurihara,C., Matsuyama,T., Kagaoka,I., Kasukawa,T., Koya,S., Kuroki,A., Miyazaki,R., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saico,H., Sakai,C., Sakai,K., Shiraishi,T., Sano,H., Sasaki,D., Shiba,K., Shinagawa,A., Tagami,M., Takeda,Y., Tagawa,A., Takanashi,F., Takaku,A., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)

REFERENCE 4 The RIKEN Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
AUTHORS Nature 420, 563-573 (2002)  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs (bases 1 to 2052)  
JOURNAL 6 (bases 1 to 2052)

REFERENCE 5 The RIKEN Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
AUTHORS Fukuda,S., Furuno,M., Haga,K., Hashizume,W., Hayashizaki,Y., Hayashida,K., Hiramoto,T., Hiraoka,T., Kasukawa,T., Horii,F., Imotani,K., Ishii,Y., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Katoch,H., Kawai,J., Kurihara,C., Matsuyama,T., Kagaoka,I., Kasukawa,T., Koya,S., Kuroki,A., Miyazaki,R., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saico,H., Sakai,C., Sakai,K., Shiraishi,T., Sano,H., Sasaki,D., Shiba,K., Shinagawa,A., Tagami,M., Takeda,Y., Tagawa,A., Takanashi,F., Takaku,A., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-JUL-2001) Yoshihiko Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, Fax: 81-45-503-9216)

COMMENT CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics

Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.

URL:<http://genome.gsc.riken.go.jp/>

Location/Qualifiers

1. .2052  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:A630041P071"  
/db\_xref="MGI:2406859"  
/db\_xref="taxon:10090"  
/clone\_id="A630041R07"  
/tissue\_type="thymus"  
/dev\_stage="3 days neonate"

1. .2052  
/note="hypothetical PDZ domain (also known as DHR or GLGF)

containing protein (InterPro|IPR001478, evidence:

InterPro)"

ORIGIN

Query Match	92.7%	Score 20.4;	DB 11;	Length 2052;	
Best Local Similarity	95.5%	Pred. No. 7.5e+02;			
Matches	21;	Mismatches	1;	Indels	0;
				Gaps	0;

1. GTGAAAGCCGCCGTGTACTCC 22  
253 GTGAAACGCCCGTGTACTCC 232

RESULT 14

BC033019/c

LOCUS BC033019 3069 bp mRNA linear HTC 19-NOV-2003  
DEFINITION Homo sapiens hypothetical protein LOC1108987, mRNA (cDNA clone IMAGE:4821752), containing frame-shift errors.  
ACCESSION BC033019  
VERSION GI:21542573  
KEYWORDS HTC  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derg,J.G., Wagner,L., Sheppard,C.M., Schuler,G.D., Klausner,R.D., Collins,F.S., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Altschul,S.F., Zeeberg,B., Jordan,H., Moore,T., Max,S.I., Wang,J.J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Stapleton,M., Soares,M.B., Borraldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carrasco,P., Prange,C., Raha,S.S., Loqueland,N.A., Peters,G.J., Abramson,R.D., Mullphy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malet,J.A., Guaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,J.J., Hulyk,S.W., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodgers,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krywinski,M.I., Skalska,U., Smalius,D.E., Scherzer,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

FEATURES source

misc\_feature

1. .2052  
/note="hypothetical PDZ domain (also known as DHR or GLGF)

containing protein (InterPro|IPR001478, evidence:

InterPro)"

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk Email: cgabbs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

Toshiyuki and Piero Carninci (RIKEN) (LNN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNN)

DNA Sequencing by: Institute for Systems Biology

contact: amadasystemsbiology.org

Anup Madan, Jessica Fahey, Erin Heiton, Mark Ketteman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

1. .2052  
/tissue\_type="thymus"  
/clone\_id="A630041R07"  
/dev\_stage="3 days neonate"

This clone was selected for full length sequencing because it passed the following selection criteria: GenoMEScan gene prediction

This clone has the following problem: frame shifted.

Location/Qualifiers

1. .2052  
/organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"

/clone IMAGE:481752  
/tissue type="Testis"  
/clone Lib="NIH MGC\_97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

FEATURES source

ORIGIN

Query Match 92.7%; Score 20.4; DB 11; Length 3069;

Best Local Similarity 95.5%; Pred. No. 7.5e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Match No. 8.1e+02;保守性 0; 错配 1; 插入 0; 缺失 0;

Score 0.1e+00; 相似性 95.5%; 预测数 7.5e+02;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

得分 0.1e+00; 相似度 95.5%; 预测数 7.5e+02;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 21; 保守性 0; 错配 1; 插入 0; 缺失 0;

匹配数 2

/codon\_start=1  
/product="similar to Drosophila melanogaster CG2254"  
/protein\_id="PARI0257\_1"  
/db\_xref="GI:31048709"  
/translation="MSKQTPSAGMNANSNDIYNIVLIVVDDIMLIVKFWIAVEAUG  
LRPEAPLDIVSGKWLITGTGHGMGKONALOZAKLGAILLCNDNEQTNNQTVKEIKS  
SGGRAFGYVCNTRKREELIELAQXVRKEHGF"

## ORIGIN

Query	1	GTGGAAGCCGGTGTGACTC	21
Db	115	GTGGAAGCCGGTGTGACTC	135

Search completed: July 18, 2004, 14:36:39  
Job time : 276.072 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:36:30 ; Search time 457.82 Seconds  
(without alignments)  
2781.271 Million cell updates/sec

Title: US-10-051-769-2

Perfect score: 261

Sequence: 1 gatcaagggtggatcgagg.....cacctgcccatacgaaatggaa 261

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3190992 seqs , 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:  
 1: /cgna2\_6/podata/2/pubna/007\_PUBCOMB.seq: \*  
 2: /cgna2\_6/podata/2/pubna/006\_NEW\_PUB.seq: \*  
 3: /cgna2\_6/podata/2/pubna/006\_PUBCOMB.seq: \*  
 4: /cgna2\_6/podata/2/pubna/007\_NEW\_PUB.seq: \*  
 5: /cgna2\_6/podata/2/pubna/007\_NEW\_PUB.seq: \*  
 6: /cgna2\_6/podata/2/pubna/007\_PUBCOMB.seq: \*  
 7: /cgna2\_6/podata/2/pubna/008\_NEW\_PUB.seq: \*  
 8: /cgna2\_6/podata/2/pubna/008\_PUBCOMB.seq: \*  
 9: /cgna2\_6/podata/2/pubna/008\_PUBCOMB.seq: \*  
 10: /cgna2\_6/podata/2/pubna/009\_PUBCOMB.seq: \*  
 11: /cgna2\_6/podata/2/pubna/009\_PUBCOMB.seq: \*  
 12: /cgna2\_6/podata/2/pubna/009\_NEW\_PUB.seq: \*  
 13: /cgna2\_6/podata/2/pubna/009\_NEW\_PUB.seq2: \*  
 14: /cgna2\_6/podata/2/pubna/010\_PUBCOMB.seq: \*  
 15: /cgna2\_6/podata/2/pubna/010\_PUBCOMB.seq: \*  
 16: /cgna2\_6/podata/2/pubna/US10\_PUBCOMB.seq: \*  
 17: /cgna2\_6/podata/2/pubna/US10\_NEW\_PUB.seq: \*  
 18: /cgna2\_6/podata/2/pubna/US60\_NEW\_PUB.seq: \*  
 19: /cgna2\_6/podata/2/pubna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	261	100.0	261	Sequence 2, Appli
2	261	100.0	261	Sequence 2, Appli
3	261	100.0	873	Sequence 3, Appli
4	261	100.0	873	Sequence 3, Appli
5	261	100.0	960	Sequence 1, Appli
6	261	100.0	960	Sequence 1, Appli
7	261	100.0	2639	Sequence 684, Appli
8	261	100.0	3465	Sequence 9, Appli
9	261	100.0	3649	Sequence 63, Appli
10	261	100.0	3832	Sequence 7, Appli
11	261	100.0	4470	Sequence 950, App
12	261	100.0	4801	Sequence 25, Appli
13	208.2	79.8	263	Sequence 4, Appli
14	208.2	79.8	263	Sequence 4, Appli

Sequence 3629, Ap  
Sequence 1, Appli  
Sequence 16261, A  
Sequence 2665, Ap  
Sequence 1, Appli  
Sequence 6721, Ap  
Sequence 33, Appli  
Sequence 3317, Ap  
Sequence 5921, Ap  
Sequence 10103, A  
Sequence 87461, A  
Sequence 45694, A  
Sequence 11, Appli  
Sequence 2587, A  
Sequence 7375, Ap  
Sequence 7822, A  
Sequence 12263, A  
Sequence 7318, Ap  
Sequence 136, App  
Sequence 2713, Ap  
Sequence 1, Appli  
Sequence 29072, A  
Sequence 3926, Ap  
Sequence 4570, Ap  
Sequence 56422, A  
Sequence 1454, Ap  
Sequence 20268, A  
Sequence 2665, Ap  
Sequence 1139, App  
Sequence 22074, A  
Sequence 32433, A

ALIGNMENTS

RESULT 1  
US-10-051-769-2  
; Sequence 2, Application US/10051769  
; Publication No. US20030044811A1  
; GENERAL INFORMATION:  
; APPLICANT: MCKINNON, Randy D.  
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
; FILE REFERENCE: 268\_260 (RWJ-00-37)  
; CURRENT APPLICATION NUMBER: US-10-051-769  
; CURRENT FILING DATE: 2001-10-20  
; PRIOR APPLICATION NUMBER: US 60/242,160  
; PRIOR FILING DATE: 2000-10-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO: 2  
; LENGTH: 261  
; TYPE: DNA  
; ORGANISM: Homo sapiens

US-10-051-769-2

Db 181 Gcccgccgcctggccggatggctggcccttgcagggtggactacaacggggcrt 240  
 Qy 241 ccacctggccatcgactgtga 261.  
 Db 241 ccacctggccatcgactgtga 261.

**RESULT 2**  
 US-10-224-624-2  
 i Sequence 2, Application US/10224624  
 i Publication No. US20030108915A1  
 i GENERAL INFORMATION:  
 i APPLICANT: MCKINNON, Randy D.  
 i TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN  
 i FILE REFERENCE: 54704-8059 US00  
 i CURRENT APPLICATION NUMBER: US/10/224,624  
 i CURRENT FILING DATE: 2002-08-20  
 i PRIOR APPLICATION NUMBER: 60/242,160  
 i PRIOR FILING DATE: 2000-10-20  
 i PRIOR APPLICATION NUMBER: 10/051,769  
 i PRIOR FILING DATE: 2001-10-20  
 i NUMBER OF SEQ ID NOS: 9  
 i SOFTWARE: PatentIn version 3.1  
 i SEQ ID NO: 2  
 i LENGTH: 261  
 i TYPE: DNA  
 i ORGANISM: Homo Sapiens

US-10-224-624-2  
 Query Match 100.0% Score 261; DB 15; Length 261;  
 Best Local Similarity 100.0% Pred. No. 3.5e-58;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGAGGCTGCTGCAGACCAAGGGCCGGCGCTGCTGGAGGG 60  
 Db 1 GATCAAGGTGGAGGCTGCTGCAGACCAAGGGCCGGCGCTGCTGGAGGG 60  
 Qy 61 GCTGAGCTCTGGGAACTGGTCTCTGGGAGACGGTGCCTTCATCAAGCATCCGGT 120  
 Db 61 GCTGAGCTCTGGGAACTGGTCTCTGGGAGACGGTGCCTTCATCAAGCATCCGGT 120  
 Qy 121 CGTCGCGGCAAGTCGCCCCGGGAACTGGGAACTGGGAACTGGGAGGGCGT 180  
 Db 121 CGTCGCGGCAAGTCGCCCCGGGAACTGGGAACTGGGAACTGGGAGGGCGT 180  
 Qy 181 GCGCGCCCTGCTGGCCGGAGGCTGGGAAACGGGGTT 240  
 Db 181 GCGCGCCCTGCTGGCCGGAGGCTGGGAAACGGGGTT 240  
 Qy 241 CCACCTGGCCATCGACGTGA 261  
 Db 241 CCACCTGGCCATCGACGTGA 261

**RESULT 3**  
 US-10-051-769-3  
 i Sequence 3, Application US/10051769  
 i Publication No. US20030044811A1  
 i GENERAL INFORMATION:  
 i APPLICANT: MCKINNON, Randy D.  
 i TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
 i FILE REFERENCE: 2681-260 (RMJ-00-37)  
 i CURRENT APPLICATION NUMBER: US/10/051,769  
 i CURRENT FILING DATE: 2001-10-20  
 i PRIOR APPLICATION NUMBER: US 60/242,160  
 i PRIOR FILING DATE: 2000-10-20  
 i NUMBER OF SEQ ID NOS: 6  
 i SOFTWARE: PatentIn version 3.1  
 i SEQ ID NO: 3  
 i LENGTH: 873  
 i TYPE: DNA  
 i ORGANISM: Homo Sapiens

Query Match 100.0% Score 261; DB 15; Length 873;  
 Best Local Similarity 100.0% Pred. No. 3e-58;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGAGGCTGCTGCAGACCAAGGGCCGGCGCTGCTGGAGGG 60  
 Db 366 GATCAAGGTGGAGGCTGCTGCAGACCAAGGGCCGGCGCTGCTGGAGGG 425  
 Qy 61 GCTGAGCTCTGGGAACTGGTCTCTGGGAGACGGTGCCTTCATCAAGCATCCGGT 120  
 Db 426 GCTGAGCTCTGGGAACTGGTCTCTGGGAGACGGTGCCTTCATCAAGCATCCGGT 485  
 Qy 121 CGTCGCGGCAAGTCGCTGGCCGGAGGGCGT 180  
 Db 486 CGTCGCGGCAAGTCGCTGGCCGGAGGGCGT 545  
 Qy 181 GCGCGCCCTGCTGGCCGGAGGCTGGGAACTGGGAGGGCGT 240  
 Db 546 GCGCGCCCTGCTGGCCGGAGGCTGGGAACTGGGAGGGCGT 605  
 Qy 241 CCACCTGGCCATCGACGTGA 261  
 Db 606 CCACCTGGCCATCGACGTGA 626

US-10-051-769-3

Query Match 100.0% Score 261; DB 15; Length 873;  
 Best Local Similarity 100.0% Pred. No. 3e-58;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGAGGCTGCTGCAGACCAAGGGCCGGCGCTGCTGGAGGG 60  
 Db 366 GATCAAGGTGGAGGCTGCTGCAGACCAAGGGCCGGCGCTGCTGGAGGG 425  
 Qy 61 GCTGAGCTCTGGGAACTGGTCTCTGGGAGACGGTGCCTTCATCAAGCATCCGGT 120  
 Db 426 GCTGAGCTCTGGGAACTGGTCTCTGGGAGACGGTGCCTTCATCAAGCATCCGGT 485  
 Qy 121 CGTCGCGGCAAGTCGCTGGCCGGAGGGCGT 180  
 Db 486 CGTCGCGGCAAGTCGCTGGCCGGAGGGCGT 545  
 Qy 181 GCGCGCCCTGCTGGCCGGAGGCTGGGAACTGGGAGGGCGT 240  
 Db 546 GCGCGCCCTGCTGGCCGGAGGCTGGGAACTGGGAGGGCGT 605  
 Qy 241 CCACCTGGCCATCGACGTGA 261  
 Db 606 CCACCTGGCCATCGACGTGA 626

RESULT 4

US-10-224-624-3  
 i Sequence 3, Application US/10224624  
 i Publication No. US20030108915A1  
 i GENERAL INFORMATION:  
 i APPLICANT: MCKINNON, Randall D.  
 i TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN  
 i FILE REFERENCE: 54704-8059-US00  
 i CURRENT APPLICATION NUMBER: US/10/224,624  
 i CURRENT FILING DATE: 2002-08-20  
 i PRIOR APPLICATION NUMBER: 60/242,160  
 i PRIOR FILING DATE: 2000-10-20  
 i PRIORITY NUMBER: 60/242,160  
 i PRIORITY NUMBER: 10/051,769  
 i NUMBER OF SEQ ID NOS: 9  
 i SOFTWARE: PatentIn version 3.1  
 i SEQ ID NO: 3  
 i LENGTH: 873  
 i TYPE: DNA  
 i ORGANISM: Homo sapiens

Query Match 100.0% Score 261; DB 15; Length 873;  
 Best Local Similarity 100.0% Pred. No. 3e-58;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGAGGCTGCTGCAGACCAAGGGCCGGCGCTGCTGGAGGG 60  
 Db 366 GATCAAGGTGGAGGCTGCTGCAGACCAAGGGCCGGCGCTGCTGGAGGG 425  
 Qy 61 GCTGAGCTCTGGGAACTGGTCTCTGGGAGACGGTGCCTTCATCAAGCATCCGGT 120  
 Db 426 GCTGAGCTCTGGGAACTGGTCTCTGGGAGACGGTGCCTTCATCAAGCATCCGGT 485  
 Qy 121 CGTCGCGGCAAGTCGCTGGCCGGAGGGCGT 180  
 Db 486 CGTCGCGGCAAGTCGCTGGCCGGAGGGCGT 545  
 Qy 181 GCGCGCCCTGCTGGCCGGAGGCTGGGAACTGGGAGGGCGT 240  
 Db 546 GCGCGCCCTGCTGGCCGGAGGCTGGGAACTGGGAGGGCGT 605  
 Qy 241 CCACCTGGCCATCGACGTGA 261  
 Db 606 CCACCTGGCCATCGACGTGA 626

RESULT 5  
 US-10-051-769-1  
 ; Sequence 1, Application US/10051769  
 ; Publication No. US200301044811A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCKINNON, Randy D.  
 ; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
 ; FILE REFERENCE: 268/260 (RMJ-00-37)  
 ; CURRENT APPLICATION NUMBER: US/10/051,769  
 ; CURRENT FILING DATE: 2001-10-20  
 ; PRIOR APPLICATION NUMBER: US 60/242,160  
 ; PRIOR FILING DATE: 2000-10-20  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 1  
 ; LENGTH: 960  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapiens  
 US-10-051-769-1

Query Match 100.0%; Score 261; DB 15; Length 960;  
 Best Local Similarity 100.0%; Pred. No. 3e-58; Indels 0; Gaps 0;  
 Matches 261; Conservative 0; Mismatches 0;

Qy 1 GATCAAGTGAATTCCGAGGAGTCGCTCGACCCAAGGGGGGGCCCTGTTGGAGGG 60  
 Db 396 GATCAAAGTGTGAGTTGAGGAGCTGGCTCGACCCAAGGGGGGGCCCTGTTGGAGGG 455  
 Qy 61 GCTGAGGCCCTGGGAGAGCTGTTCCTGGGAGAGCTGCCTCATCAAGCATCCGCT 120  
 Db 456 GCTGAGGCCCTGGGAGAGCTGTTCCTGGGAGAGCTGCCTCATCAAGCATCCGCT 515  
 Qy 121 CGTGCGGCTAGTCGCTGCCCTGGACCCGATGGGGAGGGCT 180  
 Db 516 CGTGCGGCTAGTCGCTGCCCTGGACCCGATGGGGAGGGCT 575  
 Qy 181 GCCCGCCGCTGCCGAGGAGCTGGCCCTCGAGGGAGGGCTACAAGGGGGCTT 240  
 Db 576 GCCGGCCGCTGCCGAGGAGCTGGCCCTCGAGGGAGGGCTACAAGGGGGCTT 635  
 Qy 241 CCACTGGCATCGAGTGA 261  
 Db 636 CCACTGGCATCGAGTGA 656

RESULT 7  
 US-10-104-047-684  
 ; Sequence 684, Application US/10104047  
 ; Publication No. US2003023639A1  
 ; GENERAL INFORMATION:  
 ; TITLE OF INVENTION: NO. US2003023639A1 full length cDNA  
 ; FILE REFERENCE: H-A0105  
 ; CURRENT APPLICATION NUMBER: US/10/104,047  
 ; CURRENT FILING DATE: 2002-03-25  
 ; PRIOR APPLICATION NUMBER:  
 ; PRIORITY FILING DATE:  
 ; NUMBER OF SEQ ID NOS: 4096  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO: 684  
 ; LENGTH: 2639  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-104-047-684

Query Match 100.0%; Score 261; DB 16; Length 2639;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-58; Indels 0; Gaps 0;  
 Matches 261; Conservative 0; Mismatches 0;

Qy 1 GATCAAGTGGAASTTCAGGGAGCTGCTCGAGGAGCTGCTCGAGCAAGGGGGGGCTGTTGGAGGG 60  
 Db 579 GATCAAAGTGGAASTTCAGGGAGCTGCTCGAGGAGCTGCTCGAGCAAGGGGGGGCTGTTGGAGGG 638  
 Qy 61 GCTGAGGCCCTGGGAGAGCTGTTCCTGGGAGAGCTGCCTCATCAAGCATCCGCT 120  
 Db 639 GCTGAGGCCCTGGGAGAGCTGTTCCTGGGAGAGCTGCCTCATCAAGCATCCGCT 698  
 Qy 121 CGTGCGGCTAGTCGCTGCCCTGGACCCGATGGGGAGGGCT 180  
 Db 699 CGTGCGGCTAGTCGCTGCCCTGGACCCGATGGGGAGGGCT 758  
 Qy 181 GCCGGCCGCTGCCGAGGAGCTGGCCCTCGAGGGAGGGCTACAAGGGGGCTT 240  
 Db 759 GCCGGCCGCTGCCGAGGAGCTGGCCCTCGAGGGAGGGCTACAAGGGGGCTT 818  
 Qy 241 CCACTGGCATCGAGTGA 261  
 Db 819 CCACTGGCATCGAGTGA 839

RESULT 8  
 US-10-224-624-1  
 ; Sequence 9, Application US/10224624  
 ; Publication No. US2003010891A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCKINNON, Randall D.  
 ; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten  
 ; CURRENT APPLICATION NUMBER: US/10/224,624  
 ; CURRENT FILING DATE: 2002-08-20  
 ; PRIOR APPLICATION NUMBER: 60/242,160  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 54704,8059.US00  
 ; PRIOR FILING DATE: 2001-10-20  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 1  
 ; LENGTH: 960  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapiens  
 US-10-224-624-1

Query Match 100.0%; Score 261; DB 15; Length 960;  
 Best Local Similarity 100.0%; Pred. No. 3e-58; Indels 0; Gaps 0;  
 Matches 261; Conservative 0; Mismatches 0;

Qy 1 GATCAAGTGGAGTTGAGGAGCTGCTGAGCAAGGAGCTGCTGAGGGCT 60

CURRENT APPLICATION NUMBER: US/10/224, 624  
 PRIORITY FILING DATE: 2002-08-20  
 PRIORITY APPLICATION NUMBER: 60/242, 160  
 PRIORITY FILING DATE: 2000-10-20  
 PRIORITY APPLICATION NUMBER: 10/051, 769  
 PRIORITY FILING DATE: 2001-10-20  
 NUMBER OF SEQ ID NOS: 9  
 SEQ ID NO: 9  
 LENGTH: 3465  
 ORGANISM: Homo sapiens  
 US-10-224-624-9

Query Match 100.0%; Score 261; DB 15; Length 3465;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-58;  
 Matches 261; Conservative 0; Gaps 0;

Qy 1 GATCAAGCTGGAGTTGAGGACTTGCGTCAGACCAAGACCCGGCCCTGTGGGG 60  
 Db 366 GATCAAGCTGGAGTTGAGGACTTGCGTCAGACCAAGACCCGGCCCTGTGGGG 425  
 Qy 61 GCTGAGCTGCGGGACGTTGCTGGGAGAAGCTGCTCATAGACCATCCGCT 120  
 Db 426 GCTGAGCTGCGGGACGTTGCTGGGAGAAGCTGCTCATAGACCATCCGCT 485  
 Qy 121 CCGTGGCCAGTGTGCTGCCCTGCCATGGGGATGCCATGGGGAGGGCT 180  
 Db 486 CGTGCGGCAGTGTGCTGCCCTGCCATGGGGATGCCCTGAAGGGAGGGCT 545  
 Qy 181 GCGCGCCCTGCCCCGGAGGAGCTGGCCCTGAGGGGAGGCTGAGCTAACAAGGGGGCTT 240  
 Db 546 GCGGCCCCCTGCCCTGGGAGGAGCTGGCCCTGAGGGGAGGCTT 605  
 Qy 241 CCACCTGGCCATGACCTGGA 261  
 Db 606 CCACCTGGCCATGACCTGGA 626

## RESULT 9

US-10-112-944-63  
 Sequence 63, Application US/10112944  
 Publication No. US20040048249A1

PUBLIC INFORMATION:  
 APPLICANT: Tang, Y. Tom  
 APPLICANT: Yang, Yonghong  
 APPLICANT: Weng, Gezhi  
 APPLICANT: Zhang, Jie  
 APPLICANT: Ren, Feiyun  
 APPLICANT: Xue, Aidong J.  
 APPLICANT: Wang, Jian-Rui  
 APPLICANT: Wehrman, Tom  
 APPLICANT: Ghosh, Malabika  
 APPLICANT: Wang, Dunrui  
 APPLICANT: Zhao, Qing A.  
 APPLICANT: Wang, Zhiwei

TITLE OF INVENTION: No. US20040048249A1 Nucleic Acids and

FILE REFERENCE: 805A

CURRENT APPLICATION NUMBER: US/10/112, 944  
 CURRENT FILING DATE: 2002-03-28

PRIOR APPLICATION NUMBER: US 09/488, 725  
 PRIOR FILING DATE: 2000-01-21

PRIOR APPLICATION NUMBER: US 09/491, 404  
 PRIOR FILING DATE: 2000-01-25

PRIOR APPLICATION NUMBER: US 09/515, 126  
 PRIOR FILING DATE: 2000-02-03

PRIOR APPLICATION NUMBER: US 09/519, 705  
 PRIOR FILING DATE: 2000-03-07

PRIOR APPLICATION NUMBER: US 09/540, 217  
 PRIOR FILING DATE: 2000-03-31

PRIOR APPLICATION NUMBER: US 09/552, 929  
 PRIOR FILING DATE: 2000-04-18  
 PRIORITY APPLICATION NUMBER: US 09/577, 408  
 PRIORITY FILING DATE: 2000-05-18  
 NUMBER OF SEQ ID NOS: 9245  
 SOFTWARE: pt\_Fl\_genes Version 5.0  
 SEQ ID NO: 63  
 LENGTH: 3649  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1) .. (3462)  
 US-10-112-944-63

Query Match 100.0%; Score 261; DB 13; Length 3649;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-58;  
 Matches 261; Conservative 0; Gaps 0;

Qy 1 GATCAAAGTGGACTTCGAGGAGCTGCGTCAGACCAAGACGGCGGCCCTGTGGGG 60  
 Db 366 GATCAAAGTGGACTTCGAGGAGCTGCGTCAGACCAAGACGGCGGCCCTGTGGGG 425  
 Qy 61 GCTGAGCTGCGGGACGTTGCTGGGAGAAGCTGCTCATAGACCATCCGCT 120  
 Db 426 GCTGAGCTGCGGGACGTTGCTGGGAGAAGCTGCTCATAGACCATCCGCT 485  
 Qy 121 CCGTGGCCAGTGTGCTGCCCTGCCATGGGGATGCCATGGGGAGGGCT 180  
 Db 486 CGTGC GG CAGTGTGCTGCCCTGCCATGGGGATGCCCTGAAGGGAGGGCT 545  
 Qy 181 GCGCGCCCTGCCCCGGAGGAGCTGGCCCTGAGGGGAGGCTGAGCTAACAAGGGGGCTT 240  
 Db 546 GCGGCCCCCTGCCCTGGGAGGAGCTGGCCCTGAGGGGAGGCTT 605  
 Qy 241 CCACCTGGCCATGACCTGGA 261  
 Db 606 CCACCTGGCCATGACCTGGA 626

## RESULT 10

US-10-224-624-7

Sequence 7, Application US/10224624  
 Publication No. US20030108915A1  
 GENERAL INFORMATION:  
 APPLICANT: MCKINNON, Randall D.  
 TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN  
 FILE REFERENCE: 54704 8059 US00  
 CURRENT APPLICATION NUMBER: US/10/224, 624  
 PRIOR APPLICATION NUMBER: 60/242, 160  
 PRIOR FILING DATE: 2002-08-20  
 PRIOR APPLICATION NUMBER: 60/242, 160  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 10/051, 769  
 PRIOR FILING DATE: 2001-10-20  
 NUMBER OF SEQ ID NOS: 9  
 SOFTWARE: Patentin version 3.1  
 SEQ ID NO: 7  
 LENGTH: 3832

TYPE: DNA  
 ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (178) .. (3639)

OTHER INFORMATION:  
 US-10-224-624-7

Query Match 100.0%; Score 261; DB 15; Length 3832;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-58;  
 Matches 261; Conservative 0; Gaps 0;

Qy 1 GATCAAGTGGACTTCGAGGAGCTGCGTCAGACCAAGACGGCGGCCCTGTGGGG 60  
 Db 543 GATCAAGTGGACTTCGAGGAGCTGCGTCAGACCAAGACGGCGGCCCTGTGGGG 602

RESULT 11  
 US-10-276-774-950  
 ; Sequence 950, Application US-10-276-774-950  
 ; Publication No. US20040053245A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc.  
 ; TITLE OF INVENTION: No. US20040053245A1 el Nucleic Acids and Polypeptides  
 ; FILE REFERENCE: 21272-030  
 ; CURRENT APPLICATION NUMBER: US-10/276-774  
 ; CURRENT FILING DATE: 2002-11-18  
 ; PRIOR APPLICATION NUMBER: 09-560,875  
 ; PRIOR FILING DATE: 2000-04-27  
 ; PRIOR APPLICATION NUMBER: 09-496,914  
 ; PRIOR FILING DATE: 2000-02-03  
 ; NUMBER OF SEQ ID NOS: 2700  
 ; SOFTWARE: Custom  
 ; SEQ ID NO: 950  
 ; LENGTH: 4470  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-276-774-950

Query Match 100.0%; Score 261; DB 13; Length 4470;  
 Best Local Similarity 100.0%; Prod. No. 2.5e-58; Mismatches 0; Indels 0; Gaps 0;

RESULT 13  
 US-10-051-769-4  
 ; Sequence 4, Application US-10051769-4  
 ; Publication No. US2003004481A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCKINNON, Randy D.  
 ; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
 ; FILE REFERENCE: 268/260 (RWU-00-37)  
 ; CURRENT APPLICATION NUMBER: US-10/051,769  
 ; CURRENT FILING DATE: 2001-10-20  
 ; PRIOR APPLICATION NUMBER: US 60/242,160  
 ; PRIOR FILING DATE: 2000-10-20  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 4  
 ; LENGTH: 263  
 ; TYPE: DNA  
 ; ORGANISM: Sprague Dawley rat  
 ; US-10-051-769-4

Query Match 79.8%; Score 208.2; DB 15; Length 263;  
 Best Local Similarity 87.4%; Prod. No. 1.4e-44; Mismatches 33; Indels 0; Gaps 0;

RESULT 12  
 US-10-336-603A-25  
 ; Sequence 25, Application US-10336603A  
 ; Publication No. US20040072997A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alsobrook et al.  
 ; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

RESULT 14  
 US 10-224-624-4  
 i Sequence 4, Application US/10224624  
 i Publication No. US20030108915A1  
 i GENERAL INFORMATION  
 i APPLICANT: MCKINNON, Randall D.  
 i TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten  
 i FILE REFERENCE: 54704-8059.US00  
 i CURRENT APPLICATION NUMBER: US/10/224,624  
 i CURRENT FILING DATE: 2002-08-20  
 i PRIOR APPLICATION NUMBER: 60/242,160  
 i PRIOR FILING DATE: 2000-10-20  
 i NUMBER OF SEQ ID NOS: 9  
 i SOFTWARE: PatentIn version 3.1  
 i SEQ ID NO: 4  
 i LENGTH: 263  
 i TYPE: DNA  
 i ORGANISM: Homo sapiens  
 US-10-224-624-4

Query Match Score 79.8%; Pred. No. 1.4e-44; Length 263;  
 Best Local Similarity 87.4%; 0; Mismatches 33; Indels 0; Gaps 0;

Db 1 GATCAAGTGGAGTTCAAGGACTGTAGACCAAAGGCGGGCTGAGCTTCAACCTGGCATCGAGC 257  
 Db 1 GATCAGTTGAGTTCAAGGACTGTAGACCAAAGGCGGGCTGAGCTTCAACCTGGCATCGAGC 3061

Query Match Score 208.2%; Pred. No. 1.4e-44; Length 263;  
 Best Local Similarity 91.4%; 0; Mismatches 33; Indels 0; Gaps 0;

Db 1 GATCAAGTGGAGTTCAAGGACTGTAGACCAAAGGCGGGCTGAGCTTCAACCTGGCATCGAGC 60

Query Match Score 19.6%; Pred. No. 0.0029; Length 11058;  
 Best Local Similarity 51.2%; 0; Mismatches 113; Indels 0; Gaps 0;

Db 26 CTCGAGACCAAGAGCCGGGGCTCTGGAGGGCTGAGCTGGGGACCTGGTTCCTG 85  
 Db 2830 CTGCCTCCACCTGGGCCGGCTCCCGTAGTACATGGTCCGGCGCTTGTCGCC 2889

Query Match Score 86.6%; Pred. No. 1.4e-44; Length 11058;  
 Best Local Similarity 51.2%; 0; Mismatches 113; Indels 0; Gaps 0;

Db 86 GGGAGAGGGTCCCTCATAGAACCATCGGCTCTGGGCCAGTCGTGCCCTCGGC 145  
 Db 2890 CTCGACGGCTGCCCTGACCCAAAGGCAAAGTCACAGCGGCCCTGCCGCC 2949

Query Match Score 146.0%; Pred. No. 1.4e-44; Length 11058;  
 Best Local Similarity 51.2%; 0; Mismatches 113; Indels 0; Gaps 0;

Db 146 ACCGGGAGCCGATGGCCGATGGCCGATGGGGAGGGCTGCCGCCAGGAGCTG 205  
 Db 2950 GCGAGACGCCCTGCCGCCACCTGGCCGACCTCGGCCGAGACGCCGCCGG 3009

Query Match Score 206.0%; Pred. No. 1.4e-44; Length 11058;  
 Best Local Similarity 51.2%; 0; Mismatches 113; Indels 0; Gaps 0;

Db 206 GCCTCGAGGGAGGGTGGAGTACAAGGGGGCTGCCCTCACCTGGCATCGAGC 257  
 Db 3010 ATCGCCAGGTGGCCGACCTCGGCCGACCTCGGCCGAGTCGGTGGAGG 3061

Search completed: July 18, 2004, 14:47:31  
 Job time : 463.82 secs

RESULT 15  
 US 10-156-761-3629  
 i Sequence 3629, Application US/10156761  
 i Publication No. US20030119018A1  
 i GENERAL INFORMATION:  
 i APPLICANT: OMURA, SATOSHI  
 i APPLICANT: IKEDA, HARUO  
 i APPLICANT: ISHIKAWA, JUN  
 i APPLICANT: HORIKAWA, HIROSHI  
 i APPLICANT: SHIBA, TADAYOSHI  
 i APPLICANT: SAKAKI, YOSHITOKI

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:31:29 ; Search time 435.57 Seconds  
(without alignments)  
2545.580 Million cell updates/sec

Title: US-10-051-769-2  
Perfect score: 261  
Sequence: 1 gatcaagggtggagttcgagg.....caccctggccatcgacgttggaa 261

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 3373863 seqs, 212499041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :	No.	Score	Match	Length	DB	ID	Description
	1	261	100.0	261	7	A BX95679	Abx95679 Human Gli
	2	261	100.0	261	9	A AD61560	Ad61560 Human Gli
	3	261	100.0	873	7	A BX95680	Abx95680 Human Gli
	4	261	100.0	960	7	A BX95678	Abx95678 Human Gli
	5	261	100.0	960	9	A AD61559	Ad61559 Human Gli
	6	261	100.0	1092	9	A DE07789	Ade07789 Novel cod
	7	261	100.0	2539	9	A BD62530	Adb62530 Human cDN
	8	261	100.0	3832	9	A AD61564	Ad61564 Human Gli
	9	261	100.0	4450	4	A A K51608	Aak51608 Human Pol
	10	261	100.0	4410	4	A BA09174	Aba09174 Human sec
	11	261	100.0	4470	4	A AK52592	Aak52592 Human Pol
	12	261	100.0	4470	9	A DE09891	Ade09891 Novel DNA
	13	261	100.0	4801	9	A DD49052	Add49052 Human NOV
	14	208.2	79.8	263	7	A BX95681	Abx95681 Rat Gli
	15	208.2	79.8	263	9	A AD61561	Ad61561 Human Gli
	16	169.8	65.1	1732	6	A BQ16934	Abq16934 Oligonucle
	17	169.8	65.1	1732	6	A BQ16935	Abq16935 Oligonucle
	18	137	52.5	1732	6	A BQ16937	Abq16937 Oligonucle
	19	137	52.5	1732	6	A BQ16936	Abq16936 Oligonucle
	20	51.6	19.8	6436	2	A AX86965	Aax86965 DNA conta
	21	48.6	18.6	4476	6	A BQ91125	Abq91125 M. capill
	22	47.8	18.3	2382	9	A DC30095	Adc30095 Human nov
	23	47.8	18.3	2710	6	A BN85324	Abn85324 Human cyt

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	261	100.0	261	7	A BX95679	Abx95679 Human Gli	
2	261	100.0	261	9	A AD61560	Ad61560 Human Gli	
3	261	100.0	873	7	A BX95680	Abx95680 Human Gli	
4	261	100.0	960	7	A BX95678	Abx95678 Human Gli	
5	261	100.0	960	9	A AD61559	Ad61559 Human Gli	
6	261	100.0	1092	9	A DE07789	Ade07789 Novel cod	
7	261	100.0	2539	9	A BD62530	Abd62530 Human cDN	
8	261	100.0	3832	9	A AD61564	Ad61564 Human Gli	
9	261	100.0	4450	4	A A K51608	Aak51608 Human Pol	
10	261	100.0	4410	4	A BA09174	Aba09174 Human sec	
11	261	100.0	4470	4	A AK52592	Aak52592 Human Pol	
12	261	100.0	4470	9	A DE09891	Ade09891 Novel DNA	
13	261	100.0	4801	9	A DD49052	Add49052 Human NOV	
14	208.2	79.8	263	7	A BX95681	Abx95681 Rat Gli	
15	208.2	79.8	263	9	A AD61561	Ad61561 Human Gli	
16	169.8	65.1	1732	6	A BQ16934	Abq16934 Oligonucle	
17	169.8	65.1	1732	6	A BQ16935	Abq16935 Oligonucle	
18	137	52.5	1732	6	A BQ16937	Abq16937 Oligonucle	
19	137	52.5	1732	6	A BQ16936	Abq16936 Oligonucle	
20	51.6	19.8	6436	2	A AX86965	Aax86965 DNA conta	
21	48.6	18.6	4476	6	A BQ91125	Abq91125 M. capill	
22	47.8	18.3	2382	9	A DC30095	Adc30095 Human nov	
23	47.8	18.3	2710	6	A BN85324	Abn85324 Human cyt	

#### RESULT 1

ID	ABX95679	standard; cDNA; 261 BP.
XX	AC	ABX95679;
XX	XX	30-JUN-2003 (first entry)
DE	XX	Human GLITEN expressed sequence tag.
XX	XX	Human; ss; EST; chromosome 10q25; glioblastoma multiforme; brain cancer; KW GLITEN; GMB; expressed sequence tag.
XX	OS	Homo sapiens.
XX	XX	US200304811-A1.
PD	XX	06-MAR-2003.
PP	XX	20-OCT-2001; 2001US-000511769.
PR	XX	20-OCT-2000; 2000US-0242160P.
XX	XX	(UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
PA	XX	McKinnon RD;
PI	XX	DR WPI; 2003-418995/39.
XX	XX	New isolated nucleic acid representing a gene product associated with GLITEN, useful as a molecular marker for tumor cell identification and classification.
PS	XX	Claim 1; Page 7; 11pp; English.
CC	CC	The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GLITEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control

#### ALIGNMENTS

sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into Glioblastoma Multiforme, and a kit for use in detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into Glioblastoma Multiforme (a type of brain cancer). The EST useful as molecular marker for tumour cell identification and for diagnosing or identifying candidates at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is the human GLTEN expressed sequence tag

Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 7; Length 261;

Best Local Similarity 100.0%; Pred. No. 1.6e-40;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GATCAAGGTGAGTTGAGAAGCTGTGGAGACGACCAAGAACGCCGEGCGCCCTGCTGGAGGG 60

Qy 1 GATCAAGGTGAGTTGAGAAGCTGTGGAGACGACCAAGAACGCCGEGCGCCCTGCTGGAGGG 60

Db 1 GATCAAGGTGAGTTGAGAAGCTGTGGAGACGACCAAGAACGCCGEGCGCCCTGCTGGAGGG 60

Qy 61 GCTGAGCTGGGGAGCTGTTCTGGGGAGACGGTGCCGACGGAGACGGTCCCTTCATAAAGCCTCGCT 120

Db 61 GCTGAGCTGGGGAGCTGTTCTGGGGAGACGGTGCCGACGGAGACGGTCCCTTCATAAAGCCTCGCT 120

Qy 121 CGTGCCTGGCCAGTCGTCGCCCTGGGAGCCACGGGGAGGGCCGATGCCCCTGAAGGGAGGGCT 180

Db 121 CGTGCCTGGCCAGTCGTCGCCCTGGGAGCCACGGGGAGGGCCGATGCCCCTGAAGGGAGGGCT 180

Qy 181 GCCCGCGCTGCCCCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCT 240

Db 181 GCCCGCGCTGCCCCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCT 240

Qy 241 CCACCTGGCCATGAGCTGGAA 261

Db 241 CCACCTGGCCATGAGCTGGAA 261

RESULT 3  
ABX95680 standard; DNA: 873 BP.

ID ABX95680  
XX AC ABX95680;

XX DT 30-JUN-2003 (first entry)

XX DE Human GLTEN open reading frame.

XX KW Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;

XX KW GLTEN; GMB.

XX OS Homo sapiens.

XX FH Key  
XX CDS 1..873  
XX FT /\*tag= a  
XX FT /partial  
XX FT /product= "GLTEN"  
XX FT /note= "No stop codon shown"

XX PD US2003044811-A1.

XX PN 20-OCT-2002; 2002US-00224624.

XX PR 06-MAR-2003.

XX PR 20-OCT-2001; 2001US-00051769.

XX PA 20-OCT-2000; 2000US-0242160P.

XX PA (UXNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX PI McKinnon RD;

XX DR WPI: 2003-810848/76.

XX PT Novel glioblastoma multiforme associated protein GLTEN useful for

PT treating glioblastoma multiforme and diagnosing brain cancer.

XX XX  
XX DR WPI: 2003-418995/39.  
XX PT New isolated nucleic acid representing a gene product associated with

PS Claim 38; Page 11; OPP; English.

XX The present invention provides novel glioblastoma multiforme (GBM) associated protein GLTEN useful for treating glioblastoma multiforme and diagnosing brain cancer. The invention is useful in brain cancer therapy, treatment and diagnosis. The invention is also useful in gene therapy. The present sequence is human GLTEN EST clone cDNA. This EST clone is used in the exemplification of the invention.

XX Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 9; Length 261;  
Best Local Similarity 100.0%; Pred. No. 1.6e-40;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGAGTTGAGAAGCTGTGGAGACGACCAAGAACGCCGEGCGCCCTGCTGGAGGG 60

Db 1 GATCAAGGTGAGTTGAGAAGCTGTGGAGACGACCAAGAACGCCGEGCGCCCTGCTGGAGGG 60

Qy 61 GCTGAGCTGGGGAGCTGTTCTGGGGAGACGGTGCCGACGGAGACGGTCCCTTCATAAAGCCTCGCT 120

Db 61 GCTGAGCTGGGGAGCTGTTCTGGGGAGACGGTGCCGACGGAGACGGTCCCTTCATAAAGCCTCGCT 120

Qy 121 CGTGCCTGGCCAGTCGTCGCCCTGGGAGCCACGGGGAGGGCCGATGCCCTGAAGGGAGGGCT 180

Db 121 CGTGCCTGGCCAGTCGTCGCCCTGGGAGCCACGGGGAGGGCCGATGCCCTGAAGGGAGGGCT 180

Qy 181 GCCCGCGCTGCCCCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCT 240

Db 181 GCCCGCGCTGCCCCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCTGGGGAGGGCT 240

Qy 241 CCACCTGGCCATGAGCTGGAA 261

Db 241 CCACCTGGCCATGAGCTGGAA 261

RESULT 3  
ABX95680 standard; DNA: 873 BP.

ID ABX95680  
XX AC ABX95680;

XX DT 30-JUN-2003 (first entry)

XX DE Human GLTEN open reading frame.

XX KW Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;

XX KW GLTEN; GMB.

XX OS Homo sapiens.

XX FH Key  
XX CDS 1..873  
XX FT /\*tag= a  
XX FT /partial  
XX FT /product= "GLTEN"  
XX FT /note= "No stop codon shown"

XX PD US2003044811-A1.

XX PN 20-OCT-2002; 2002US-00224624.

XX PR 06-MAR-2003.

XX PR 20-OCT-2001; 2001US-00051769.

XX PA 20-OCT-2000; 2000US-0242160P.

XX PA (UXNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX PI McKinnon RD;

XX DR WPI: 2003-810848/76.

XX PT Novel glioblastoma multiforme associated protein GLTEN useful for

PT treating glioblastoma multiforme and diagnosing brain cancer.

XX XX  
XX DR WPI: 2003-418995/39.  
XX PT New isolated nucleic acid representing a gene product associated with

PT Glioblastoma Multiforme, designated as GLITEN, useful as a molecular  
 PT marker for tumor cell identification and classification.

PS Disclosure; Page 7; 11pp; English.

XX The invention relates to an isolated nucleic acid representing a gene  
 CC product associated with Glioblastoma Multiforme, designated as GLITEN,  
 CC comprising the human EST (expressed sequence tag) appearing as ABX95678,  
 CC or a sequence that hybridises under stringent conditions to the EST, or  
 CC its complement. Also included are a probe for use in identifying a  
 CC patient at risk for progression into the malignant phenotype, comprising  
 CC the nucleic acids detailed above, detecting whether a patient is at risk  
 CC for progression into Glioblastoma Multiforme (comprising: (a) providing a  
 CC sample from a patient; (b) adding the probe to the sample or performing a  
 CC analysis using the primers appearing as ABX95682 and ABX95683; (c)  
 CC analysing levels of mRNA bound with the probe; and (d) treating a control  
 CC sample to assess the level of mRNA in a control sample, where the  
 CC presence of increased levels of mRNA expression in the sample in an  
 CC amount higher than the control sample indicates risk for progression into  
 CC Glioblastoma Multiforme), and a kit for use in detecting whether a  
 CC patient is at risk for progression into Glioblastoma Multiforme  
 CC (comprising the probe or primers). The probe or the kit is useful for  
 CC detecting whether a patient is at risk for progression into Glioblastoma  
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker  
 CC for tumour cell identification and classification and for diagnosing or  
 CC identifying candidates at risk for progression into a malignant phenotype  
 CC especially in brain cancer therapy, treatment and diagnosis. The present  
 XX sequence is the open reading frame from the GLITEN gene

Sequence 873 BP; 137 A; 304 C; 279 G; 153 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 7; Length 873;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-40;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAAGTGGAGTTGAGGAGCTGCTGAGCACGAGCAAGCAGCCCTGTGAGGG 60  
 Db 366 GATCAAAGTGGAGTTGAGGAGCTGCTGAGCACGAGCAAGCAGCCCTGTGAGGG 425  
 Qy 61 GCTGAGGCTGGAGACGCTGTCGGGAGAAGGTGCCCTCATAGACCATCCGCT 120  
 Db 426 GCTGAGGCTGGAGACGCTGTCGGGAGAAGGTGCCCTCATAGACCATCCGCT 485  
 Qy 121 CCGTGGCCCAAGTGGCTGGCCACCGGGGATGGCCGATGCCCTGAAGGGAGGGCT 180  
 Db 486 CCGTGGCCCAAGTGGCTGGCCACCGGGGATGGCCGATGCCCTGAAGGGAGGGCT 545  
 Qy 181 GCGCGCGCTGCCCGAGGAGCTGGGCTTCGAGGGAGGTGAGTACAACGGGGCTT 240  
 Db 546 GCGCGCGCTGCCCGAGGAGCTGGGCTTCGAGGGAGGTGAGTACAACGGGGCTT 605

Qy 241 CCACCTGGCATAGCTGTTGGA 261  
 Db 606 CCACCTGGCATAGCTGTTGGA 626

RESULT 4  
 ABX95678 ID ABX95678 standard; DNA; 960 BP.  
 XX AC ABX95678;  
 XX DT 30-JUN-2003 (first entry)

XX DE Human gene encoding GLITEN.

XX KW Human; ds; gene; chromosome 10q25; glioblastoma multiforme; brain cancer; GLITEN; GMB.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers  
 FT CDS 30..906

PT /\*tag= a  
 PT /product= "GLITEN"  
 XX PN US2003044811-A1.

XX PD 06-MAR-2003.

XX PR 20-OCT-2001; 2001US-00051769.  
 XX PR 20-OCT-2000; 2000US-0242160P.

XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX PI McKinnon RD;  
 XX DR WPI; 2003-418995/39.

XX New isolated nucleic acid representing a gene product associated with  
 PT Glioblastoma Multiforme, designated as GLITEN, useful as a molecular  
 PT marker for tumor cell identification and classification.  
 XX PS Example 2; Fig 1; 11pp; English.

XX The invention relates to an isolated nucleic acid representing a gene  
 CC product associated with Glioblastoma Multiforme, designated as GLITEN,  
 CC comprising the human EST (expressed sequence tag) appearing as ABX95678,  
 CC or a sequence that hybridises under stringent conditions to the EST, or  
 CC its complement. Also included are a probe for use in identifying a  
 CC patient at risk for progression into the malignant phenotype, comprising  
 CC the nucleic acids detailed above, detecting whether a patient is at risk  
 CC for progression into Glioblastoma Multiforme (comprising: (a) providing a  
 CC sample from a patient; (b) adding the probe to the sample or performing a  
 CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)  
 CC analysing levels of mRNA bound with the probe; and (d) treating a control  
 CC sample to assess the level of mRNA in a control sample, where the  
 CC presence of increased levels of mRNA expression in the sample in an  
 CC amount higher than the control sample indicates risk for progression into  
 CC Glioblastoma Multiforme), and a kit for use in detecting whether a  
 CC patient is at risk for progression into Glioblastoma Multiforme  
 CC (comprising the probe or primers). The probe or the kit is useful for  
 CC detecting whether a patient is at risk for progression into Glioblastoma  
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker  
 CC for tumour cell identification and classification and for diagnosing or  
 CC identifying candidates at risk for progression into a malignant phenotype  
 CC especially in brain cancer therapy, treatment and diagnosis. The Present  
 CC sequence is a region of human chromosome 10 (q25) containing the GLITEN  
 CC gene

XX Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;

XX Query Match 100.0%; Score 261; DB 7; Length 960;  
 XX Best Local Similarity 100.0%; Pred. No. 1.5e-40;  
 XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGTGGAGTTGAGGAGCTGCTGAGCACGAGCAAGCAGCCCTGTGAGGG 60  
 Db 396 GATCAAGTGGAGTTGAGGAGCTGCTGAGCACGAGCAAGCAGCCCTGTGAGGG 455

Qy 61 GCTGAGGCTGGAGACGCTGTCGGGAGAAGGTGCCCTCATAGACCATCCGCT 120

Db 456 GCTGAGGCTGGAGACGCTGTCGGGAGAAGGTGCCCTCATAGACCATCCGCT 515

Qy 121 CCGTGGCCCAAGTGGCTGGCCACCGGGGATGGCCGATGCCCTGAAGGGAGGGCT 180

Db 486 CCGTGGCCCAAGTGGCTGGCCACCGGGGATGGCCGATGCCCTGAAGGGAGGGCT 575

Qy 181 GCGCGCGCTGCCCGAGGAGCTGGGCTTCGAGGGAGGTGAGTACAACGGGGCTT 240

Db 536 GCGCGCGCTGCCCGAGGAGCTGGGCTTCGAGGGAGGTGAGTACAACGGGGCTT 635

Qy 241 CCACCTGGCATAGCTGTTGGA 261

Db 636 CCACCTGGCATAGCTGTTGGA 656

RESULT 5

ID AAD61559 standard; cDNA; 960 BP.

XX AAD61559;

AC AAD61559;

DT 15-JAN-2004 (first entry)

XX Human Gliten partial cDNA.

DE Human; glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis; gene therapy; chromosome 10; gene; ss.

KW Unidentified.

XX XX WO2003054152-A2.

OS Homo sapiens.

XX PN US2003108915-A1.

XX PD 12-JUN-2003.

XX PF 20-AUG-2002; 2002US-00224624.

XX PR 20-OCT-2000; 2000US-0242160P.

PR 20-OCT-2001; 2001US-00051769.

PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

PI McKinnon RD;

XX DR 2003-810848/76.

XX WPI: 2003-810848/76.

PT Novel glioblastoma multiforme associated protein Gliten useful for treating glioblastoma multiforme and diagnosing brain cancer.

XX PS Example 2; Fig 1; Opp: English.

XX The present invention provides novel glioblastoma multiforme (GBM) associated protein Gliten useful for treating glioblastoma multiforme and diagnosing brain cancer. The invention is useful in brain cancer therapy, treatment and diagnosis. The invention is also useful in gene therapy. The present sequence is human Gliten partial cDNA. The human Gliten gene is located on chromosome 10. This partial cDNA is used in the exemplification of the invention.

CC Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;

CC Query Match 100.0%; Score 261; DB 9; Length 960;

CC Best Local Similarity 100.0%; Pred. No. 1.5e-40;

CC Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC Qy 1 GATCAAGGTGAGTTCGAGAACGACGGCTGTTCTGGCGAGCACGGT 120

Db 396 GATCAGGTGAGTTCGAGAACGACGGCTGTTCTGGCGAGCACGGT 455

Qy 61 GCTGACCTCTGGGAAGCTGTTCTGGCGAGCACGGCTCATCAAGCCATTGGCT 120

Db 456 GCTGACCTCTGGGAAGCTGTTCTGGCGAGCACGGCTCATCAAGCCATTGGCT 515

Qy 121 CGTGCGGCAGACTCGTGGCTCGGCCACGGAGGCCATGGCTGAAGGGAGGGCT 180

Db 516 CGTGCGGCAGACTCGTGGCTCGGCCACGGAGGCCATGGCTGAAGGGAGGGCT 575

Qy 181 GCCCGCGCCCTGGCCGAGGAGCTGGCTGAGGGAGCCATGGCTGAAGGGAGGGCT 240

Db 576 GCCCGCGCCCTGGCCGAGGAGCTGGCTGAGGGAGCCATGGCTGAAGGGAGGGCT 635

Qy 241 CCACCTGGCCATGAGTGG 261

Db 636 CCACCTGGCCATGAGTGG 656

---

Query Match 100.0%; Score 261; DB 9; Length 1092;

Best Local Similarity 100.0%; Pred. No. 1.5e-40;

Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 366 GATCAAGGTGAGTTCGAGAACGACGGCTGTTCTGGCGAGCACGGCT 425

Qy 61 GCTGACCTCTGGGAAGCTGTTCTGGCGAGCACGGCTCATCAAGCCATTGGCT 120

Db 426 GCTGACCTCTGGGAAGCTGTTCTGGCGAGCACGGCTCATCAAGCCATTGGCT 485

Qy 121 CGTGCGGCAGACTCGTGGCTCGGCCACGGAGGCCATGGCTGAAGGGAGGGCT 180

Db 486 CGTGCGGCAGACTCGTGGCTCGGCCACGGAGCTGGCTGAAGGGAGGGCT 545

Qy 181 GCCCGCGCCCTGGCCGAGGAGCTGGCTGAGGGAGCCATGGCTGAAGGGAGGGCT 240

RESULT 6

Db 546 GCGCGCCCTGCCGAGGAGCTGGCTTCAGGGGGAGGTACACGGGGCTT 605  
 Qy 241 CAACTGGCATCGACCTGGA 261  
 Db 606 CCACCTGGCATCGACCTGGA 626

RESULT 7  
 ADB62530 standard; cDNA; 2639 BP.  
 XX ADB62530;  
 AC ADB62530;  
 DT 04-DEC-2003 (first entry)

XX Human cDNA encoding clone HHDP20088160.

XX Human; ss: gene; pharmaceutical; diagnostic; gene therapy;  
 KW tissue regeneration; cell regeneration; membrane protein;  
 KW signal transduction-related protein; transcription-related protein;  
 KW osteoporosis; neurological disease; cancer; tumour.

XX Homo sapiens.

FH Key Location/Qualifiers  
 FT CDS 214..2639  
 FT /tag= a  
 FT /partial=  
 FT /product= "Clone HHDP20088160 protein"  
 PN EP1308459-A2.  
 XX PD 07-MAY-2003.  
 XX PP 28-MAR-2002; 2002EP-00007401.  
 XX PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JUN-2002; 2002US-00350978.  
 XX PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikata T, Otsuka M, Nagahari K, Masuno Y,  
 XX WPI: 2003-450961/43.  
 DR P-PSDB; ADB64500.  
 XX P-PDB; ADB64500.

PT New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.

XX PS Claim 1; Page: 222pp; English.

XX The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesising the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis,

CC neurological diseases, cancer, tumours. The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a cDNA of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

SQ Sequence 2639 BP; 657 A; 702 C; 719 G; 561 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 9;  
 Best Local Similarity 100.0%; Pred. No. 1..4e-40;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GATCAAGCTGAGACTGAGGAGCTGAGCAAGACGCGGGCTGCTGAGGG 60  
 DB 579 GATCAAGCTGAGACTGAGGAGCTGAGCAAGACGCGGGCTGCTGAGGG 638

Oy 61 GCTGAGCTGCGGGAGCTGCTGCGAGAGCGCTCCCTCATCAAGACCATCGGGCT 120  
 DB 639 GCTGAGCTGCGGGAGCTGCTGCGAGAGCGCTCCCTCATCAAGACCATCGGGCT 698

Oy 121 CGTGCGGCCAGTCAGTCGCTCGCTCGCTCGCTCGCTGCAGCCGAGGGAGGCT 180  
 DB 699 CGTGCGGCCAGTCAGTCGCTCGCTCGCTCGCTGCAGCCGAGGGAGGCT 758

Oy 181 GCGCGCCCTCCCGAGGAGCTGCCTTCAGGGAGGTGAGTACAAGGGCCTT 240  
 DB 759 GCGCGCCCTCCCGAGGAGCTGCCTTCAGGGAGGTGAGTACAAGGGCCTT 818

Oy 241 CCACCTGCCATCGACTGGGA 261  
 DB 819 CCACCTGCCATCGACTGGGA 839

RESULT 8  
 AAD61564  
 ID AAD61564 standard; cDNA; 3832 BP.  
 XX AAD61564;  
 AC AAD61564;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE Human Gliten cDNA.  
 XX Human glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;  
 KW gene therapy; chromosome 10; gene; ss.  
 XX OS Homo sapiens.

FH Key  
 CDS 178..3642  
 FT /tag= a  
 FT /product= "Human Gliten protein"  
 FT /note= "The CDS is specifically claimed in claim 1"

FT primer\_bind  
 FT FT 1011..1032  
 FT /tag= b  
 FT /bound\_moiety= "5' primer #843"  
 FT primer\_bind  
 FT FT 1405..1429  
 FT /tag= c  
 FT /bound\_moiety= "5' primer #1405"  
 FT primer\_bind  
 FT FT 2083..2108  
 FT /tag= d  
 FT /bound\_moiety= "5' primer #2083"  
 FT primer\_bind  
 FT FT 2356..2378  
 FT /tag= e  
 FT /bound\_moiety= "5' primer #2355"  
 FT primer\_bind  
 FT FT 3625..3650  
 FT /tag= f  
 FT /bound\_moiety= "3' primer #3636"  
 XX US200108915-A1.  
 XX PN  
 PD 12-JUN-2003.  
 XX

PF 20-AUG-2002; 2002US-00224624.  
 XX PR 03-FBB-2000; 2000US-00446914.  
 PR 20-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00538075.  
 PR 19-JUL-2000; 2000US-00630325.  
 PR 01-SEP-2000; 2000US-00634936.  
 PR 15-SEP-2000; 2000US-00633561.  
 PR 20-OCT-2000; 2000US-00633325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX (HYSE-) HYSEQ INC.  
 XX PA  
 DR (P-PSDB; ABW00758.  
 XX PT Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Wang J, Ren F, Chen R, Wang ZW;  
 PT Ma Y, Zhao QA, Wang D, Zhang J, Weihran T, Goodrich R;  
 XX PS AJ, Yang Y,  
 XX WPI; 2001-476283/51.  
 DR P-PSDB; AAM78475.  
 CC Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 CC in diagnosis and gene therapy.  
 XX PT  
 XX PCT 2001-476283/51.  
 CC Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 CC in diagnosis and gene therapy.  
 XX PT  
 XX PCT 2001-476283/51.  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AM78323-AM80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The vaccines or  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haemopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC antiinhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX SQ Sequence 3832 BP; 1090 A; 926 C; 957 G; 859 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 261; DB 9; Length 3832;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-40;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GATCAAGGTGAGTTGAGGAAGCCTGCAGACCCAGACGGCGCCGCTGCTGAGGG 60  
 Db 543 GATCAAGGTGAGTTGAGGAAGCCTGCAGACCCAGACGGCGCCGCTGCTGAGGG 602  
 Qy 61 GCTGAGCCCTGGGGAGCTGTTCCCTGGAGACGTGCTCCATTCAAGACCATCGGCT 120  
 Db 603 GCTGAGCCCTGGGGAGCTGTTCCCTGGGGAGACGTGCTCCATTCAAGACCATCGGCT 662  
 Qy 121 CGTGCAGCAGTCGTGCCCTGGGACCCGGGACGCCGATGCCCTGAACGGGAGCGCT 180  
 Db 663 CGTGCAGCAGTCGTGCCCTGGGACCCGGGACGCCGATGCCCTGAACGGGAGCGCT 722  
 QY 181 GCCCCCGCCCTGGGGAGCTGAGGAGCTGGCTTGAGAGGAGCTGTAACGCGGGCT 240  
 Db 723 GCCCCGGCCCTGGGGAGCTGAGGAGCTGGCTTGAGAGGAGCTGTAACGCGGGCT 782  
 QY 241 CCACCTGGCATTCGACGTGGA 261  
 Db 783 CCACCTGGCATTCGACGTGGA 803  
 RESULT 9  
 AAK51608  
 ID AAK51608 standard; cDNA; 4450 BP.  
 XX AC  
 AC AAK51608;  
 XX DT 06-NOV-2001 (first entry)  
 XX DE Human polynucleotide SEQ ID NO 153.  
 XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haemopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157190-A2.  
 XX PD 09-AUG-2001.  
 XX PF 05-FEB-2001; 2001WO-US004098.  
 RESULT 10  
 ABA09174  
 ID ABA09174 standard; cDNA; 4470 BP.  
 XX AC  
 AC ABA09174;  
 XX DT 11-JAN-2002 (first entry)



PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX  
 PS Claim 1: Page 4496-4497; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM8323-AAM8302) that exhibit activity eliciting to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haemopoiesis regulating activities,  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activation/inhibition activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM8020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX  
 Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;  
 SQ Query Match 100.0%; Score 261; DB 4; Length 4470;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-40;  
 Matches 261; Conservative 0; Mismatches 0; Gaps 0;  
 Qy 1 GATCAAGTGGAGCTTCAGGAGCTGCTCAGACCAAGGCCGGCCCTCTGGAGGG 60  
 Db 366 GATCAAGTGGAGCTTCAGGAGCTGCTCAGACCAAGGCCGGCCCTCTGGAGGG 425  
 Qy 61 GCTGAGCTGGGAGCCTGTTCTGGGAGAGCGGTCCCTCATCAAACATCCGGT 120  
 Db 426 GCTGAGCTGGGAGCCTGTTCTGGGAGAGCGGTCCCTCATCAAACATCCGGT 485  
 Qy 121 CGTGCAGCAGTGTGCTCGCCACCGGAGCCGATGCCCTGAAGGGAGGGCT 180  
 Db 486 CGTGCAGCAGTGTGCTCGCCACCGGAGCCGATGCCCTGAAGGGAGGGCT 545  
 Qy 181 GCGCGCCCTGCGGAGGAGCTGGCTTCAGGGGAGGGAGGTAAACGGGGCTT 240  
 Db 546 GCGCGCCCTGCGGAGGAGCTGGCTTCAGGGGAGGGAGGTGGACTAACGGGGCTT 605  
 Qy 241 CCACCTGCCCATGACGTGGA 261  
 Db 606 CCACCTGCCCATGACGTGGA 626

RESULT 1.2  
 ADE09891  
 ID ADE09891 standard; DNA; 4470 BP.  
 XX  
 AC ADE09891;  
 XX DT 29-JAN-2004 (first entry)  
 XX  
 DE Novel DNA-related contig nucleotide sequence #613.  
 XX  
 KW novel gene; novel protein; tissue marker; molecular weight marker;  
 KW chromosome marker; genetic disorder; contig; ds.  
 XX OS Unidentified.  
 PN WO2003054152-A2.  
 XX DT 03-JUL-2003.  
 XX PR 10-DEC-2002; 2002WO-US039555.  
 XX PR 10-DEC-2001; 2001US-0339719P.  
 PR 11-DEC-2001; 2001US-0339453P.  
 PR 14-MAR-2002; 2002US-0365091P.  
 PR 14-MAR-2002; 2002US-0365384P.  
 PR 12-APR-2002; 2002US-0372381P.  
 PR 12-APR-2002; 2002US-0372615P.  
 PR 22-APR-2002; 2002US-00128558.  
 PR 24-APR-2002; 2002US-0376045P.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
 PI Ghosh M, Xue AJ, Wehrman T, Zhou P, Drmanac RT, Wang Z;  
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
 XX DR WPI; 2003-569235/53.  
 XX PT New polynucleotides, useful for expressing recombinant proteins for  
 PT analysis, characterization or therapeutic use, or as markers for tissues  
 PT in which the corresponding protein is preferentially expressed.  
 XX DR Disclosure; SEQ ID NO 2435; 117pp; English.  
 XX  
 CC The invention comprises the amino acid and coding sequences of novel  
 CC proteins. The DNA and protein sequences of the invention are useful as:  
 CC markers for tissues in which the corresponding protein is preferentially expressed; as molecular weight markers on gels; as chromosome markers or  
 CC tags; to identify chromosomes or to map related gene positions; and to  
 CC compare with endogenous DNA sequences in patients to identify potential  
 CC generic disorders. The present DNA sequence was used in the  
 CC exemplification of the invention.  
 XX  
 SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 261; DB 4; Length 4470;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-40;  
 Matches 261; Conservative 0; Mismatches 0; Gaps 0;  
 Qy 1 GATCAAGTGGAGCTTCAGGAGCTGCTCAGACCAAGGCCGGCCCTCTGGAGGG 60  
 Db 366 GATCAAGTGGAGCTTCAGGAGCTGCTCAGACCAAGGCCGGCCCTCTGGAGGG 425  
 Qy 61 GCTGAGCTGGGAGCCTGTTCTGGGAGAGCGGTCCCTCATCAAACATCCGGT 120  
 Db 426 GCTGAGCTGGGAGCCTGTTCTGGGAGAGCGGTCCCTCATCAAACATCCGGT 485  
 Qy 121 CGTGCAGCAGTGTGCTCGCCACCGGAGCCGATGCCCTGAAGGGAGGGCT 180  
 Db 486 CGTGCAGCAGTGTGCTCGCCACCGGAGCCGATGCCCTGAAGGGAGGGCT 545  
 Qy 181 GCGCGCCCTGCGGAGGAGCTGGCTTCAGGGGAGGGAGGTAAACGGGGCTT 240  
 Db 546 GCGCGCCCTGCGGAGGAGCTGGCTTCAGGGGAGGGAGGTGGACTAACGGGGCTT 605  
 Qy 241 CCACCTGCCCATGACGTGGA 261  
 Db 606 CCACCTGCCCATGACGTGGA 626  
 XX  
 DE Human NOV6a coding sequence, SEQ ID 25.  
 XX  
 KW Antidiabetic; anorectic; cardiotonic; hypotensive; antiarteriosclerotic;  
 KW viricide; antibacterial; fungicide; protozoacide; nootropic;  
 KW neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;  
 KW antiarthritic; antiinflammatory; dermatologic; antiasthmatic;  
 KW antihypertensive; gene therapy; NOV protein; metabolic disorder; diabetes;  
 KW obesity; viral infection; bacterial infection; fungal infection;  
 KW helminthic infection; protozoal infection; anorexia; cancer;  
 KW cardiovascular disease; hypertension; atherosclerosis;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;

KW inflammatory skin disorder; asthma; dyslipidemia; human; gene; ds.  
 XX Homo sapiens.  
 XX WO2003060149-A2.  
 PN 24-JUL-2003.  
 XX 06-JAN-2003; 2003WO-US000252.  
 PR 04-JAN-2002; 2002US-0345222P.  
 PR 14-JAN-2002; 2002US-0348633P.  
 PR 16-JAN-2002; 2002US-0349182P.  
 PR 17-JAN-2002; 2002US-0349733P.  
 PR 18-JAN-2002; 2002US-0350263P.  
 PR 24-JAN-2002; 2002US-0351977P.  
 PR 08-MAY-2002; 2002US-0383758P.  
 PR 05-JUN-2002; 2002US-0385969P.  
 PR 11-JUN-2002; 2002US-0387834P.  
 PR 17-JUL-2002; 2002US-0396407P.  
 PR 30-SEP-2002; 2002US-0415115P.  
 PR 03-JAN-2003; 2003US-00336603.  
 XX (CURA-) CORAGEN CORP.  
 PA DR; 2003-507288/55.  
 XX DR; ADD49053.  
 PS Claim 20; Page 126-127; 311pp; English.  
 XX The present invention relates to novel NOV proteins and their coding sequences (ADD49028-Add4911). The proteins and coding sequences are useful in the manufacture of a medicament for a syndrome associated with a human disease, preferably a NOV-associated disorder such as metabolic disorders, diabetes, obesity, infections diseases (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer, cardiovascular diseases (hypertension, atherosclerosis), neurodegenerative disorders, Alzheimer's disease, neurofibrotic disorders (osteoarthritis), hematopoietic disorders, epilepsy, immune disorders (osteoarthritis), inflammatory skin disorders, asthma and various dyslipidemias. The coding sequences and proteins may also be used as targets for the identification of small molecules that modulate or inhibit e.g. neurogenesis, cell differentiation, cell proliferation, haematopoiesis, wound healing and angiogenesis, in gene therapy, in generation of antibodies that bind immunospecifically to NOV substances for use in therapeutic or diagnostic methods.  
 CC Sequence 4801 BP; 1358 A; 1037 C; 1142 G; 1264 T; 0 U; 0 Other;  
 CC Query Match 100.0%; Score 261; DB 9; Length 4801;  
 CC Best Local Similarity 100.0%; Pred. No 1.3e-40;  
 CC Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GATCAAGGNGGAGTCGAGGAGCTGCTGA GACCAAGAGGCCCTGCTGGAGGG 60  
 DB 543 GATCAAGGNGGAGTCGAGGAGCTGCTGA GACCAAGAGGCCCTGCTGGAGGG 602  
 QY 61 GCTGAGCCCTGGGAGCTGTCTCGGGAGACGGTGCCTCATCACACATCCGCT 120  
 DB 603 GCTGAGCCCTGGGAGCTGTCTCGGGAGACGGTGCCTCATCACACATCCGCT 662  
 SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

QY 121 CGTGCCGCCAGTCGTCGCCCTGGGAGGCCATGCCCTGAGGGAGGGCT 180  
 DB 663 CGTGCCGCCAGTCGTCGCCCTGGGAGGCCATGCCCTGAGGGAGGGCT 722  
 QY 181 GCGCGCGCCAGCCGGAGGAGCTGCCCTGAGGGAGGGACTAACAGGGCTT 240  
 DB 723 GCGCGCGCCAGCCGGAGGAGCTGCCCTGAGGGAGGGACTAACAGGGCTT 782  
 QY 241 CCACCTGGCCATCGACCTGGA 261  
 DB 783 CCACCTGGCCATCGACCTGGA 803

RESULT 14  
 ID ABX95681 standard; cDNA; 263 BP.  
 XX ABX95681;  
 AC AC  
 XX 30-JUN-2003 (first entry)  
 DT DT  
 DE Rat Gliten expressed sequence tag.  
 XX Rat; ss; EST; glioblastoma multiforme; brain cancer; Gliten; GMB;  
 KW expressed sequence tag.  
 XX  
 XX Rattus Norvegicus.  
 OS OS  
 XX US2003044811-A1.  
 PN PN  
 XX PD 06-MAR-2003.  
 XX PF 20-OCT-2001; 2001US-00051769.  
 XX PR 20-OCT-2000; 2000US-0242160P.  
 PA (Dyne-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 XX PI McKinnon RD;  
 XX DR; 2003-418995/39.  
 PT New isolated nucleic acid representing a gene product associated with glioblastoma multiforme, designated as Gliten, useful as a molecular marker for tumor cell identification and classification.  
 XX Example 2; Page 4; 11PP; English.  
 XX The invention relates to an isolated nucleic acid representing a gene product associated with glioblastoma multiforme, designated as Gliten, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into glioblastoma multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control sample to assess the level of mRNA in a control sample, where the presence of increased levels of mRNA expression in the sample in an amount higher than the control sample indicates risk for progression into glioblastoma multiforme), and a kit for use in detecting whether a patient is at risk for progression into glioblastoma multiforme (comprising the probe or primers). The probe or the kit is useful for detecting whether a patient is at risk for progression into glioblastoma multiforme (a type of brain cancer). The EST is useful as a molecular marker for tumour cell identification and classification and for diagnosing or identifying candidates at risk for progression into a malignant phenotype especially in brain cancer therapy, treatment and diagnosis. The present sequence is a Rat Gliten expressed sequence tag  
 SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

Query Match Similarity 79.8%; Score 208.2; DB 7; Length 263;  
 Best Local Similarity 87.4%; Pred. No. 1.4e-30; Mismatches 33; Indels 0; Gaps 0;  
 Matches 228; Conservative 0; Gaps 0;

Qy 1 GATCAAAGTGGAGTTGAGGAAGTCCTGCCTGAGACCAAGACGGCGGGCCCTGCTGGAGGG 60  
 Db 1 GATCAAAGTGGAGTTGAGGAAGTCCTGCCTGAGACCAAGACGGCGGGCCCTGCTGGAGGG 60

Qy 61 GCTGAGCTGCGGAGCTGGTTCCTGGGAGAGGTGCCCTCATAGACCATCCGGCT 120  
 Db 61 GCTGAGCTGCGGAGCTGGTTCCTGGGAGAGGTGCCCTCATAGACCATCCGGCT 120

Qy 121 CGTGCGGCAAGTGGCCTCGGCACTGGGAGAGGGGATGGCCCTGAGGGAGGGCT 180  
 Db 121 CGTGCGGCAAGTGGCCTCGGCACTGGGAGAGGGGATGGCCCTGAGGGAGGGCT 180

Qy 181 GCGCGCCGCTGCCCGAGGGATGCCATGGCGATGCCATGGCGATGCCATGGCGCT 180  
 Db 121 GCTGCGGCACTGGCTCGGCACTGGGAGAGGGGATGGCCCTGAGGGAGGGCT 180

Qy 181 GCGCGCCGCTGCCCGAGGGATGCCATGGCGATGCCATGGCGATGCCATGGCGCT 240  
 Db 181 GCGCGCCGCTGCCCGAGGGGACTGGCTTGAGCGAGGTACAACGGGGCTT 240

Qy 241 CCACCTGGCCATGGCTGGCAACGGGGCTT 261  
 Db 241 CCACCTGGCCATGGCTGGCAACGGGGCTT 261

Qy 241 CCACCTGGCCATGGCTGGCAACGGGGCTT 261  
 Db 241 CCACCTGGCCATGGCTGGCAACGGGGCTT 261

Search completed: July 18, 2004, 12:45:02  
 Job time : 438.57 secs

## RESULT 15

AAD61561  
 ID AAD61561 standard; cDNA; 263 BP.  
 XX  
 AC AAD61561;  
 XX  
 DT 15-JAN-2004 (first entry)

DE Human Gliten EST clone cDNA. #2.  
 XX Human; glioblastoma multiforme; GBM; Gliten; brain cancer; diagnosis;  
 KW gene therapy; expressed sequence tag; ss.  
 XX Homo sapiens.  
 OS  
 PN US2003108915-A1.  
 PD 12-JUN-2003.

PP 20-AUG-2002; 2002US-00224624.  
 XX  
 PR 20-OCT-2000; 2000US-0242160P.  
 PR 20-OCT-2001; 2001US-00051769.  
 XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 PI McKinnon RD;  
 XX  
 DR WPI; 2003-810848/76.

PT Novel glioblastoma multiforme associated protein Gliten useful for  
 XX treating glioblastoma multiforme and diagnosing brain cancer.  
 PS Example 2; Page 8; Opp; English.  
 XX  
 PT The present invention provides novel glioblastoma multiforme (GBM)  
 CC associated protein Gliten useful for treating glioblastoma multiforme and  
 CC diagnosing brain cancer. The invention is useful in brain cancer therapy,  
 CC treatment and diagnosis. The invention is also useful in gene therapy.  
 CC The present sequence is human Gliten EST clone cDNA. This EST clone is  
 CC used in the exemplification of the invention  
 XX Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

SQ Query Match Similarity 79.8%; Score 208.2; DB 9; Length 263;  
 Best Local Similarity 87.4%; Pred. No. 1.4e-30; Mismatches 33; Indels 0; Gaps 0;  
 Matches 228; Conservative 0; Gaps 0;



Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FABESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=MR2&t2=MR2-BN0364-  
280800-005-f07&t3=2000-08-28&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 10  
High quality sequence stop: 581.  
Location/Qualifiers  
1. .583  
source  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="BN0364"  
/note="Organ: breast-normal; Vector: puc18; Site 1: Smal;  
Site 2: SmaI; A mani-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
Profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
ORIGIN  
Query Match 100.0% Score 261; DB 10; Length 583;  
Best Local Similarity 100.0% Pred. No. 1.9e-32;  
Matches 261; Conservative 0; Mismatches 0; Gaps 0;  
Qy 1 GATCAAGGTGAAAGTTCGAGGAAGCTGTCTGAGACCAAGAACCGCCGGCTGGTGGAGG 60  
Db 482 GATCAGGTGAAAGTTCGAGGAAGCTGTCTGAGACCAAGAACCGCCGGCTGGTGGAGG 423  
Qy 61 GCTGAGGCTGGGGAGCTGGTCTGGGAGAGCTGGTCTGGAGACCAATCGGGCT 120  
Db 422 GCTGAGGCTGGGGAGCTGGTCTGGAGACCAATCGGGCT 363  
Qy 121 CGTGCAGCACTCGTGCCTCGGCCACCGGGAGCCGATGCCCTTGAGACGAGCTGGCT 180  
Db 362 CGTGCAGCACTCGTGCCTCGGCCACCGGGAGCCGATGCCCTTGAGACGAGCTGGCT 303  
Qy 181 GCCGCGCCCTGCCCGAAGAGCTGGCTGCCCTTGAGGGAGGTGGATCAAACGGGGCTT 240  
Db 302 GCCCGCGCCCTGCCCGAAGAGCTGGCTGCCCTTGAGGGAGGTGGATCAAACGGGGCTT 243  
Qy 241 CCACCTGGCCATCGACTGGAA 261  
Db 242 CCACCTGGCCATCGACTGGAA 222

RESULT 2  
BC033019 LOCUS BC033019  
DEFINITION Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone INAYKE:1821752), containing frame-shift errors.  
ACCESSION BC033019  
VERSION GI:21542573  
KEYWORDS HTC.  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 3069)  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Schuler, G.D., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schaefer, C.F., Bhat, N.K., Altschul, S.F., Zeeberg, B.R., Buetow, K.H., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Sharpsteen, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquai, L., Peters, G.J.,

Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McErlean, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Murny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Ketteman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, B.D., Dickson, M.C., Rodriguez, A.C., Grimmrod, J., Schmidt, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Stalski, U., Sali, D.E., Schnierch, A., Schein, J.E., Jones, S.J. and Marra, M.A., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
JOURNAL MEDLINE  
MEDLINE 23288257  
PUBMED 12477932  
REFERENCE 2 (bases 1 to 3069)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC). Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2599, USA  
NIH-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
http://www.systemsbiology.org  
contact: amadan@systemsbiology.org  
Anup Madan, Jessica Fahay, Erin Helton, Mark Kerteman, Anuradha Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting  
COMMENT This clone has the following qualifiers  
1. .3069  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone=IMAGE:4821752"  
/tissue type="Testis"  
/clone\_id="NIH\_MGC\_97"  
/lab\_ho="DH10B"  
/note="Vector: bluescript"  
ORIGIN  
Query Match 100.0% Score 261; DB 11; Length 3069;  
Best Local Similarity 100.0% Pred. No. 2.6e-32;  
Matches 261; Conservative 0; Mismatches 0; Gaps 0;  
Qy 1 GATCAAGGTGAAAGTTCGAGGAAGCTGTCTGAGACCAATCGGGCT 729  
Db 610 GATCAAGGTGAAAGTTCGAGGAAGCTGTCTGAGACCAATCGGGCT 669  
Qy 61 GCTGAGGCTGGGGAGCTGGTCTGGGAGAGCTGGTCTGGAGACCAATCGGGCT 120  
Db 610 GATCAAGGTGAAAGTTCGAGGAAGCTGTCTGAGACCAATCGGGCT 60  
This Clone was selected for full length sequencing because it passed the following selection criteria: GenoMeScan gene prediction  
This clone has the following problem: frame shifted.  
Location/Qualifiers  
1. .3069  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/clone=IMAGE:4821752"  
/tissue type="Testis"  
/clone\_id="NIH\_MGC\_97"  
/lab\_ho="DH10B"  
/note="Vector: bluescript"

Qy	241	CCACCTGGCATCGACGTGGA	261				RESULT 4
Db	850	CCACCTGGCATCGACGTGGA	870				BB638573
RESULT 3							BB638573
DEFINITION	BX41552	989 bp mRNA linear EST 02-MAY-2003		458 bp mRNA	linear	EST 31-AUG-2001	RIKEN full-length enriched, 3 days neonate thymus Mus
ACCESSION	BX41552	Homo sapiens HBL CELLS COT 25-NORMALIZED Homo sapiens					musculus cDNA clone A63004107 5', mRNA sequence.
VERSION	BX41552.1	GT:30339998					BB638573
SOURCE	Homo sapiens (human)						EST
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathia; Murinae; Mus.						Mus musculus (house mouse)
KEYWORDS							
REFERENCE							1 (bases 1 to 458)
AUTHORS	Arakawa,T., Carninci,P., Fukuda,S., Furutani,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Konda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nonura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasai,D., Shibusawa,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.						
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)						
JOURNAL							Unpublished (2001)
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute of Physical and Chemical Research (RIKEN) 1-7-22 Sebircho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9226 Email: genome_r@gscc.riken.go.jp/ URL:http://genome.gsc.riken.go.jp/						
FEATURES	source	/organism="Homo sapiens" /mol_type="mRNA" /cloneline="HELA CELLS COT 25-NORMALIZED" /clone="CSODK02BA06QPI2" /db_xref="taxon:9606" /db_xref="taxon:9606" /clone="CSODK02YA12" /cell_line="HELA"					
FEATURES	source	/clone="1st strand cDNA was primed with a NotI-oligo (dr) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized." /origin="Homo sapiens" /mol_type="mRNA" /cloneline="Homo sapiens HEla CELLS COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo (dr) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."					
ORIGIN							
Qy	99.2%	Score 259; DB 13; Length 989;					Query Match 99.2% Score 259; DB 13; Length 989;
Db	99.2%	Pred. No. 4.3e-32; Indels 0; Gaps 0;					Best Local Similarity 99.2%; Pred. No. 4.3e-32; Indels 0; Gaps 0;
Matches	259	Conservative 1; Mismatches 1;					Matches 259; Conservative 1; Mismatches 1;
Qy	1	GATCAAGTGGAGTTGGCTTGCTGAGCAAGAACGCGCCGCTCTGGAGGG 60					Db 633 GATCAAGTGGAGTTGGCTTGCTGAGCAAGAACGCGCCGCTCTGGAGGG 692
Db	633	GATCAAGTGGAGTTGGCTTGCTGAGCAAGAACGCGCCGCTCTGGAGGG 692					
Qy	61	GATGAGGCTGGGGGAGCTGGTCTCTGGAGGAGGCTTCATCAAGAACATCGGGCT 120					Db 693 GCTGAGCTGGGGAGCTGGTCTCTGGAGGAGGCTTCATCAAGAACATCGGGCT 752
Db	693	GCTGAGCTGGGGAGCTGGTCTCTGGAGGAGGCTTCATCAAGAACATCGGGCT 752					
Qy	121	CCTGGGGCCAGTGCTGCCCTGGCCACGGGAGGCCGATGGCCCTGAAGGGAGGGCT 180					Db 753 CCTGGGGCCAGTGCTGCCCTGGCCACGGGAGGCCGATGGCCCTGAAGGGAGGGCT 812
Db	753	CCTGGGGCCAGTGCTGCCCTGGCCACGGGAGGCCGATGGCCCTGAAGGGAGGGCT 812					
Qy	181	GCCGGCCCTGCCGAGGAGTGGCTTGAGGGAGGGCTGAGATAACGGGGCTT 240					Db 813 GCCGGCCCTGCCGAGGAGTGGCTTGAGGGAGGGCTGAGATAACGGGGCTT 872
Db	813	GCCGGCCCTGCCGAGGAGTGGCTTGAGGGAGGGCTGAGATAACGGGGCTT 872					
Qy	241	CCACCTGGACCATGACCTGGGA 261					Db 873 CCACCTGGACCATGACCTGGGA 893
Db	873	CCACCTGGACCATGACCTGGGA 893					

/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was



FEATURES	Location/Qualifiers
source	1. '930 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N-3" /db_xref="taxon:10090" /clone="IMAGE:643-9752" /tissue_type="tumor", biopsy sample" /dev_stage="5 months" /lab_host="DH10B" /clone_idb="NCI-CGAP_Mam2"
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
ORIGIN	
	Query Match Score 87.1%; Best Local Similarity 92.0%; Matches 240; Conservative 0; Mismatches 21; Indels 0; Gaps 0; Score 227.4; DB 13; Length 930;
	YQ 1 GATCAAGCTGGAGTCGAGAGCTGCTGCAACAAAGACGCCGGAGCCGCTGTGGAGGG 60 Db 335 GATCAAGTGGAAGTTGAGTCGAGCTGGCGCTGCAACAAAGCGCCGGCTGTGGAGGG 394 Qy 61 GATGAGGCTGGGGACGCTGTTCTGGCGGAAGCGGACCATCGGGCT 120 Db 395 GCTGAGATCTGGGGACCGTGTCTGGGGAAACGGTGCCTCATCAAGACCTGGCT 454 Qy 121 CGTGCGGCCAGGTGTGCCTCGGACACGGGCTGAGGCCCTGAAGGGGAGGGCT 180 Db 455 GTGTGGCCCGTGGCTGGCTGGCCACGGGAGCCGGAGCGGGACGGCT 514 Qy 181 GCGCGCCCTGCCGGAGGACTGGCTGGCCCTGAGGGAGGTACAAAGGGGCTT 240 Db 515 CGCGCACCTGCGGAGGAGTGGCTTGTAGCGAGTAGACAGCGGCTT 574 Qy 241 CGACCTGGCCATCGACCTGGGA 261 Db 575 CCACCTGCCATCGACCTGGGA 595
RESULT 7	
AK041847	AK041847 2052 bp mRNA linear HTC 19-SEP-2003 DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone-A630041P07 product:hypothetical PDZ domain (also known as DHR or GLGF) containing protein, full insert sequence.
ACCESSION AK041847	AK041847_1 GI:26088683
VERSION 1	HTC; CAP trapper. Mus musculus (house mouse)
KEYWORDS	Mammalia; Eutheria; Rodentia; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Muridae; Murinae; Mus.
ORGANISM	
REFERENCE 1	Carninci,P., and Hayashizaki,Y. High efficiency full-length cDNA cloning Title Meth. Enzymol. 303, 19-44 (1999)
AUTHORS	Itoh,M., Kono,H., Okazaki,Y., Muranatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
PUBLMED	11042174
REFERENCE 3	Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishime,T., Harada,A.,
AUTHORS	Query Match Score 227.4; DB 11; Length 2052; Best Local Similarity 92.0%; Pred. No. 5.7e-27;
REFERENCE	5. The RIKEN Consortium and the RIKEN Genome Exploration Research Group Phase II Team. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
AUTHORS	
JOURNAL	
REFERENCE 4	The RIKEN Genome Exploration Research Group Phase II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagata,N., Hiraoka,T., Hirozane,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katch,H., Kawai,J., Kohima,Y., Kondo,S., Konno,H., Kuroda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Onsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sasaki,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shiba,T., Shirakawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muranatsu,M. and Hayashizaki,Y.
COMMENT	Direct Submission Submitted (16-JUL-2001) Yoshihide HAYASHIZAKI, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 230-0045, Japan (E-mail: genome-res@gs.c.u-tokyo.ac.jp, URL: http://genome-gsc.riken.go.jp/, Tel.: +81-45-503-9216), Fax: +81-45-503-9222, URL: http://genome-gsc.riken.go.jp/, Tel.: +81-45-503-9222, URL: http://fantom.gsc.riken.go.jp/.
FEATURES	
source	1. .2055 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB: A630041P07" /db_xref="MGI:2406859" /db_xref="Ensembl:10090" /clone="A630041P07" /tissue_type="thymus" /clone_id="RIKEN full-length enriched mouse cDNA library" /dev_stage="3 days neonate" 1. .2052 /note="hypothetical PDZ domain (also known as DHR or GLGF) containing protein (InterPro IPR001478, evidence: InterPro)"
ORIGIN	



RESULT	1.0	DEFINITION	Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp/ URL: http://genome.gsc.riken.go.jp/	LOCUS	BX345030	DEFINITION	944 bp mRNA COT 25-NORMALIZED Homo sapiens	DEFINITION	01-MAY-2003		
FEATURES	source	Organism	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1. (bases 1 to 944)	ACCESSION	BX345030	VERSION	BX345030.1	GI	30313335		
FEATURES	KEYWORDS	AUTHORS	Li, W.B., Gruber,C., Jesse,J. and Polayes,D.	TITLE		JOURNAL					
FEATURES	SOURCE	ORGANISM	Homo sapiens (human)	COMMENT							
FEATURES	COMMENT	REFERENCE	Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5483.r. For more information about this cluster, see http://www.genoscope.cn.fr/cgi-bin/cluster.cgi?seq=CS2BAX16ZB10_AX27ZD11_1&cluster=5483.r.	CONTACT							
FEATURES	source	FEATURES	1. .944 /organism="Homo sapiens" /mol type="mRNA" /db_xref="taxon:9606" /clone="CS0DK02YA12" /cell_type="HELA CELLS COT 25-NORMALIZED" /cell_line="HELA"	ORIGIN							
FEATURES	source	COMMENT	/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and ECO R sites of the PCMVSPORT 6 vector. Library was normalized."	Query Match	77.8%; Score 203; DB 13; Length 944;	Best Local Similarity	95.8%; Pred. No. 4e-23;	Mismatches	7; Indels 4; Gaps 4;	RESULT	11
FEATURES	source	COMMENT	1. .437 /organism="Mus musculus" /mol type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="K430318E15" /tissue type="visual cortex" /clone_lib="RIKEN full-length enriched, visual cortex"	Match	1 GATCAAGGTGGAGTTGGAGGCTCCTGCAGCCAAAGGCCGCGCTGGCTGGGG 60	Matches	250; Conservative 0; Mismatches 0;	Qy	1 GATCAAGGTGGAGTTGGAGGCTCCTGCAGCCAAAGGCCGCGCTGGCTGGGG 60	DEFINITION	BB866050
FEATURES	source	COMMENT	1. .437 /organism="Mus musculus" /mol type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="K430318E15" /tissue type="visual cortex" /clone_lib="RIKEN full-length enriched, visual cortex"	Match	61 GATCAAGGTGGAGGCTCCTGCAGCCAAAGGCCGCGCTGGCTGGGG 700	Matches	239; Conservative 0; Mismatches 0;	Db	61 GATCAAGGTGGAGGCTCCTGCAGCCAAAGGCCGCGCTGGCTGGGG 700	DEFINITION	BB866050
FEATURES	source	COMMENT	1. .437 /organism="Mus musculus" /mol type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="K430318E15" /tissue type="visual cortex" /clone_lib="RIKEN full-length enriched, visual cortex"	Match	701 GCTGCGCCAGTCGTGCGCCCTGGAGCCACGGGGATGCCCTGAGGGGGCTT 120	Matches	239; Conservative 0; Mismatches 0;	Qy	701 GCTGCGCCAGTCGTGCGCCCTGGAGCCACGGGGATGCCCTGAGGGGGCTT 120	DEFINITION	BB866050
FEATURES	source	COMMENT	1. .437 /organism="Mus musculus" /mol type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="K430318E15" /tissue type="visual cortex" /clone_lib="RIKEN full-length enriched, visual cortex"	Match	761 CGTGGCCAGTCGTGCGCCCTGGAGCCACGGGGATGCCCTGAGGGGGCTT 180	Matches	239; Conservative 0; Mismatches 0;	Db	761 CGTGGCCAGTCGTGCGCCCTGGAGCCACGGGGATGCCCTGAGGGGGCTT 180	DEFINITION	BB866050
FEATURES	source	COMMENT	1. .437 /organism="Mus musculus" /mol type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="K430318E15" /tissue type="visual cortex" /clone_lib="RIKEN full-length enriched, visual cortex"	Match	820 GCGCCGGNCTGCCCTGGAGCTGGCC-TCGAGGGGGAGGCTGAGTC 240	Matches	239; Conservative 0; Mismatches 0;	Qy	820 GCGCCGGNCTGCCCTGGAGCTGGCC-TCGAGGGGGAGGCTGAGTC 240	DEFINITION	BB866050
FEATURES	source	COMMENT	1. .437 /organism="Mus musculus" /mol type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="K430318E15" /tissue type="visual cortex" /clone_lib="RIKEN full-length enriched, visual cortex"	Match	877 CAACCTGGCCATGGCTGGGA 897	Matches	239; Conservative 0; Mismatches 0;	Db	877 CAACCTGGCCATGGCTGGGA 897	DEFINITION	BB866050

ACCESSION	BB86050	musculus cDNA clone G431003009 5', mRNA sequence.
VERSION	BB86050.1	GI:17112260
KEYWORDS		
SOURCE	Mus musculus Mus musculus (house mouse)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathini; Muridae; Murinae; Mus.	
REFERENCE	Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouada, M., Matsuyama, T., Nakamura, M., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sozabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takanaka-Akanira, S., Tanaka, T., Tomaru, A., Toya, T., Watanuki, A., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.	
AUTHORS	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)	
TITLE	Unpublished. (2001)	
JOURNAL	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Fax: 81-45-503-9222 Email: genome.rse@gsc.riken.go.jp/ URL: http://genome.gsc.riken.go.jp/ Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. <i>Genome Res.</i> 10 (10):1617-1630 (2000) Wegi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuuwa, S., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., and Hayashizaki, Y.	
COMMENT	RIKEN integrated sequence analysis (RISA) system--B4-format sequencing pipeline with 384 multicapillary sequencer. <i>Genome Res.</i> 10 (11), 175-1771 (2000) Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. <i>Genome Res.</i> 11 (2), 281-289 (2001) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. e mouse tissues.	
FEATURES	Location/Qualifiers	
source	1. .681 .organism="Mus musculus" .mol_type="mRNA" .strain="BA1B/c" .db_xref="taxon:10090" .clone="G431003009" .clone_line="CRL-1751 WEHI 164" .clone_lib="RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA"	
ORIGIN	Query Match 39.3%; Score 102.6; DB 13; Length 730; Best Local Similarity 68.2%; Pred. No. 4.7e-07; Matches 178; Conservative 0; Mismatches 44; Indels 39; Gaps 1;	
Qy	1 GATCAAGTGGAGTTCGAGAGCTGTGAGACAAAGCGCCGCGCCTCTCATCAAAGGAAAGGTGTGAGGG 60	
Db	172 GATCAAGTGGAGTTCGAGAGCTGTGAGACAAATGACGGAAAGGTGTGAGGG 231	
Qy	61 GCTGAGCCCTCGGAGCTGCTGGGGAGACGGCTGCTGGGGCTATGGCAATGTGCTGGGGATGGGG 120	
Db	232 GCTGAGCCCTCGGAGCTGCTGGGGATGGGGCTATGGCAATGTGCTGGGGATGGGG 291	
Qy	121 CGCTGGGCAAGTCGCTGGGGACACGGGGCTGATGGGGATGGGGCTAAGGGGG 180	
Db	292 CATCGGCCGTTCTGGGGACACGGGGCTGATGGGGCTAAGGGGG 681	



Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Numazaki, R., Ohno, M., Obsato, N., Saito, R., Sakazumi, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeeda, Y., Waki, K., Watanuki, A., Muramatsu, M., and Hayashizaki, Y.  
Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

## FEATURES

source 1.  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="Raxxon:10096"  
/clone="G83002:K01"  
/tissue type="mammary gland"  
/cell\_line="RCB-0526 JyG-MC (A)"  
/clone lib="RIKEN full-length enriched, mammary gland RCB-0526 JyG-MC (A) cDNA"

## ORIGIN

Query Match 35.5%; Score 92.6; DB 13; Length 667;  
Matches 106; Conservate 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGAGACCAAAGCCGGCCCTGCTGGAGGG 60  
Db 556 GATCAAAGCTGGAGTTCTGGAGCTGCTGAGACCAAAGCCGGCCCTGCTGGAGGG 615  
Qy 61 GCTGAGGCTGGGAGCAGNTTCTGGGAGACGGTGCCTCATCAAAC 111  
Db 616 GCTGAGGCTGGGAGCAGNTTCTGGGAGACGGTGCCTCATCAAAC 665

## RESULT

AW013379 LOCUS AW013379 723 bp mRNA linear EST 10-SEP-1999  
DEFINITION p042ks Winter flounder spleen Pseudopleuronectes americanus [Caenorhabditis elegans], mRNA  
SEQUENCE clone sp042ks 5, similar to C53B4.4

## ACCESSION

AW013379

EST

Pseudopleuronectes americanus (winter flounder)

SOURCE Pseudopleuronectidae; Pleuronectidae; Pseudopleuronectes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygi; Percomorpha; Pleuronectiformes; Pleuronectoidea; Pleuronectidae; Pseudopleuronectes

REFERENCE 1 (bases 1 to 723)

AUTHORS Douglas, S.B., Gallant, J.W., Bullerwell, C.E., Wolff, C., Munholland, J. and Reith, M.E.

TITLE Winter Flounder expressed sequence tags: Establishment of an EST database and identification of novel fish genes

JOURNAL Marine Biotechnology (1999) In press  
COMMENT Marine Biology  
NRC Institute for Marine Biosciences

1411 Oxford St., Halifax, Nova Scotia, B3H 3Z1, Canada  
Tel: (902) 426-8276  
Fax: (902) 426-9413  
Email: michael.reith@nrc.ca

Seq primer: M13 Forward  
Location/Qualifiers

1. -723  
/organism="Pseudopleuronectes americanus"

/mol\_type="mRNA"  
/db\_xref="Raxxon:8265"

/clone="sp042ks"

/sex="female"  
/dev\_stage="adult"

/clone lib="Winter flounder spleen"

/note="Organ: spleen"

## ORIGIN

Query Match 34.5%; Score 90; DB 9; Length 723;  
Best Local Similarity 69.1%; Pred. No. 4.9e-05;  
Matches 123; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGAGTTCTGGAGCTGCTGAGACCAAAGCCGGCCCTGCTGGAGGG 60  
Db 482 GATCAAAGTTGGAGTTCTGGAGCTGCTGAGACCAAAGCCGGCCCTGCTGGAGGG 541  
Qy 61 GCTGAGGCTGGGGAGCTGGTTCCTGGGAACTTCCTCCGGCAATTCTGCCGCTCAAGGGGCCACCT 610  
Db 542 GCTCAAGCTCAGAGACATCTCCCTGGCAATTCTGCCGCTCAAGGGGCCACCT 601  
Qy 121 CGTGCGGCCAGTCGTCGCCACGGGGAGCCGATGGCCCTGAAGGGGGCG 178  
Db 602 CATGAAAGCCGGTCACTGAAACCTGAAAGCGCATGCGGAAGAACCTGAGGGG 659

Search completed: July 18, 2004, 14:36:32  
Job time : 3242.26 secs

```
post-processing: Listing first 45 summaries  
database : Published Applications NA:  
           1: /cgnd2_6/podata/2/pubnpna/us07 PUBCOMB.seq:*
```

/cgns\_6/PTodata/2/pubpna/PCT

No. is the number of results predicted by chance to have a score greater than or equal to the result being printed and is derived by analysis of the total score distribution.

result No.	Query				Description	
	Score	Match Length	DB	ID		
1	261	100.0	261	15	US-10-051-769-2	Sequence 2, Appli
2	261	100.0	261	15	US-10-224-624-2	Sequence 2, Appli
3	261	100.0	873	15	US-10-051-769-3	Sequence 3, Appli
4	261	100.0	873	15	US-10-224-624-3	Sequence 3, Appli
5	261	100.0	960	15	US-10-051-769-1	Sequence 1, Appli
6	261	100.0	960	15	US-10-224-624-1	Sequence 1, Appli
7	261	100.0	2639	16	US-10-104-624-684	Sequence 684, Appli
8	261	100.0	3465	15	US-10-224-624-9	Sequence 9, Appli
9	261	100.0	3649	13	US-10-112-944-63	Sequence 6, Appli
10	261	100.0	3832	15	US-10-224-624-7	Sequence 7, Appli
11	261	100.0	4470	13	US-10-276-774-950	Sequence 950, Appli
12	261	100.0	4801	12	US-10-336-603A-25	Sequence 25, Appli
13	47	18.0	3847	12	US-10-250-823-11	Sequence 11, Appli
14	43	16.5	263	15	US-10-051-769-4	Sequence 4, appli

Db 181 GCCCGCCGCTGCCCTGGAGGTGAGCTGCCCTTGAGGGAGGTGAGTAACGGGGCCT 240  
 Qy 241 CCACCTGCCATTCGAGTGGAA 261  
 Db 241 CCACCTGCCATTCGAGTGGAA 261

RESULT 2  
US-10-224-624-2

Sequence 2 Application US/10224624

Publication No. US20050108915A1

GENERAL INFORMATION:

APPLICANT: MCKINNON, Randal D.

TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN

FILE REFERENCE: 54704-8059-US00

CURRENT APPLICATION NUMBER: US/10/224,624

CURRENT FILING DATE: 2002-08-20

PRIOR APPLICATION NUMBER: 60/242,160

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 10/051,769

NUMBER OF SEQ ID NOS: 9

SOFTWARE: Patentin version 3.1

SEQ ID NO 2

LENGTH: 261

TYPE: DNA

ORGANISM: Homo Sapiens

US-10-224-624-2

Query Match 100.0%; Score 261; DB 15; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-123;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGAGTTGGAGCTGTGAGCACCAAGAACGCCGCCTGCTGGAGGG 60

Db 1 GATCAAGGTGGAGTTGGAGCTGTGAGCACCAAGAACGCCGCCTGCTGGAGGG 60

Qy 61 GCTGAGCTGGGGAGTGTCTGGAGACGAGCTGTGAGCACCAAGAACGCCGCCT 120

Db 61 GCTGAGCTGGGGAGTGTCTGGAGACGAGCTGTGAGCACCAAGAACGCCGCCT 120

Qy 121 CGTGGCGCAGCTGIGGCCCTGGCCACGGGAGGCCGATGGCCCTGAAGGGAGGGCT 180

Db 121 CGTGGCGCAGCTGIGGCCCTGGCCACGGGAGGCCGATGGCCCTGAAGGGAGGGCT 180

Qy 181 GCCCGCGCCATGCCGAGGACTGCTTGAGGGAGGTACACGGGGCT 240

Db 181 GCCCGCGCCATGCCGAGGACTGCTTGAGGGAGGTACACGGGGCT 240

Qy 241 CCACCTGCCATTCGAGTGGAA 261

Db 241 CCACCTGCCATTCGAGTGGAA 261

RESULT 3  
US-10-051-769-3

Sequence 3 Application US/10051769

Publication No. US2003044811A1

GENERAL INFORMATION:

APPLICANT: MCKINNON, Randy D.

TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION

FILE REFERENCE: 268/260 (RN-00-37)

CURRENT APPLICATION NUMBER: US/10/051,769

CURRENT FILING DATE: 2001-10-20

PRIOR APPLICATION NUMBER: US 60/242,160

PRIOR FILING DATE: 2000-10-20

NUMBER OF SEQ ID NOS: 6

SOFTWARE: Patentin version 3.1

SEQ ID NO 3

LENGTH: 873

TYPE: DNA

ORGANISM: Homo Sapiens

US-10-051-769-3

Query Match 100.0%; Score 261; DB 15; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-123;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGAGTTGGAGCTGTGAGCACCAAGAACGCCGCCTGCTGGAGGG 60

Db 366 GATCAAGGTGGAGTTGGAGCTGTGAGCACCAAGAACGCCGCCTGCTGGAGGG 425

Qy 61 GCTGAGCTGGCTGGAGCTGTGAGCACCAAGAACGCCGCCTGCTGGAGGG 120

Db 426 GCTGAGCTGGCTGGAGCTGTGAGCACCAAGAACGCCGCCTGCTGGAGGG 485

Qy 121 CGTGGCGCCAGCTGGCCCTGGCAACGGGAGGCCGATGGCCCTGAAGGGAGGGCT 180

Db 486 CGTGGCGCCAGCTGGCCCTGGCAACGGGAGGCCGATGGCCCTGAAGGGAGGGCT 545

Qy 181 GCCCGCGCCATGCCGAGGACTGCTTGAGGGAGGTACACGGGGCT 240

Db 546 GCCCGCGCCATGCCGAGGACTGCTTGAGGGAGGTACACGGGGCT 605

Qy 241 CCACCTGCCATTCGAGTGGAA 261

Db 606 CCACCTGCCATTCGAGTGGAA 626

RESULT 5  
US-10-051-769-1  
; Sequence 1, Application US/10051769  
; Publication No. US2003004481A1  
; GENERAL INFORMATION:  
; APPLICANT: MCKINNON, Randy D.  
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
; FILE REFERENCE: 268/260 (RMJ-0-3)  
; CURRENT FILING NUMBER: US/10-051-769  
; PRIORITY FILING DATE: 2001-10-20  
; PRIOR APPLICATION NUMBER: US 6,072,421,660  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 960  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-10-051-769-1

Query Match Similarity 100.0%; Score 261; DB 15; Length 960;  
Best Local Similarity 100.0%; Pred. No. 5 6e-123; Indels 0; Gaps 0;  
Matches 261; Conservative 0; Mismatches 0; Insertions 0; Deletions 0;

Qy 1 GATCAAGGTGAGCTTCGAGGAGCTGGTGTGTCAGACCAAGACGGCGGGGCGCTGCTGAGGG 60  
Db 396 GATCAAGGTGAGCTTCGAGGAGCTGGTGTGTCAGACCAAGACGGCGGGGCGCTGCTGAGGG 455  
Qy 61 GCTGAGCTGGGGAGCTGTTCTGGGGAGAGGTGCCCTCATCAAGCCATCCGGCT 120  
Db 456 GCTGAGCTGGGGAGCTGTTCTGGGGAGAGGTGCCCTCATCAAGCCATCCGGCT 515  
Qy 121 CGTGCGGCAAGTGTGCCCTGGCCAAGGGCGCATGGCTGAAGGGAGGGCT 180  
Db 516 CGTGCGGCAAGTGTGCCCTGGCCAAGGGCGCATGGCTGAAGGGAGGGCT 575  
Qy 181 GGCCTGGCCCTGGCCAGAGGTGGCTTCAGGGAGGTGAGTACAACGGGGTT 240  
Db 576 GGCCTGGCCCTGGCCAGAGGTGGCTTCAGGGAGGTGAGTACAACGGGGTT 635

Qy 241 CCACCTGGCCATGACGTGGA 261  
Db 636 CCACCTGGCCATGACGTGGA 656

RESULT 7  
US-10-104-047-684  
; Sequence 684, Application US/10104047  
; Publication No. US2003023639A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US2003023639A1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104-047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIORITY FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 684  
; LENGTH: 2639  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-684

Query Match Similarity 100.0%; Score 261; DB 16; Length 2639;  
Best Local Similarity 100.0%; Pred. No. 4.7e-123; Indels 0; Gaps 0;  
Matches 261; Conservative 0; Mismatches 0; Insertions 0; Deletions 0;

Qy 1 GATCAAGGTGGAGTTCCAGGAGCTGCCTCACACGGGGCGCTGCTGAGGG 60  
Db 579 GATCAAGGTGGAGTTCCAGGAGCTGCCTCACACGGGGCGCTGCTGAGGG 638  
Qy 61 GCTGAGCTGGGGAGCTGGTGTGCCCTGGCCAGAGGTGCCCTCATCAAGCCATCCGGCT 120  
Db 639 GCTGAGCTGGGGAGCTGGTGTGCCCTGGCCAGAGGTGCCCTCATCAAGCCATCCGGCT 698  
Qy 121 CGTGCGGCAAGTGTGCCCTGGCCAAGGGCGCATGGCTGAAGGGAGGGCT 180  
Db 659 CGTGCGGCAAGTGTGCCCTGGCCAAGGGCGCATGGCTGAAGGGAGGGCT 758  
Qy 181 GGCCTGGCCCTGGCCAGAGGTGGCTTCAGGGAGGTGAGTACAACGGGGCTT 240  
Db 759 GGCCTGGCCCTGGCCAGAGGTGGCTTCAGGGAGGTGAGTACAACGGGGCTT 818

Qy 241 CCACCTGGCCATGACGTGGA 261  
Db 819 CCACCTGGCCATGACGTGGA 839

RESULT 8  
US-10-224-624-9  
; Sequence 9, Application US/10224624  
; Publication No. US2003010891A1  
; GENERAL INFORMATION:  
; APPLICANT: MCKINNON, Randall D.  
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN  
; FILE REFERENCE: 54704-8059 US00  
; CURRENT APPLICATION NUMBER: US/10/224,624  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: 6,072,421,660  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 10-051-769  
; PRIOR FILING DATE: 2001-10-20  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 960  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-10-224-624-1

Query Match Similarity 100.0%; Score 261; DB 15; Length 960;  
Best Local Similarity 100.0%; Pred. No. 5 6e-123; Indels 0; Gaps 0;  
Matches 261; Conservative 0; Mismatches 0;

Qy 1 GATCAAGGTGAGTTCGAGGAGCTGGCTGACGACCAAGACGGGGCGCTGCTGAGGG 60

```

; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 9
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-9

Query Match 100.0%; Score: 261; DB 15; Length: 3465;
Best Local Similarity 100.0%; Pred. No. 4.5e-123; Mismatches 0; Indels 0; Gaps 0;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAGTTGAGAATGGAGCTGAGACCAAGAGCCGGCGCCGTGCTGAGGG 60
Db 366 GATCAGTTGAGAATGGAGCTGAGACCAAGAGCCGGCGCCGTGCTGAGGG 425
Qy 61 GCTGAGGCCTGGGGAGCTGGTTCCTGGGAGAACGCTGGCCCTCATCAAGACCATAAGGGCT 120
Db 426 GCTGAGGCCTGGGGAGCTGGTTCCTGGGAGAACGCTGGCCCTCATCAAGACCATAAGGGCT 485
Qy 121 CGTGCGGCCAGCTCGCTGGGCCACCGGGAGGCCGATGCCCTGAAAGGGAGSGCCT 180
Db 486 CGTGCGGCCAGCTCGCTGGGCCACCGGGAGGCCGATGCCCTGAAAGGGAGSGCCT 545
Qy 181 GCGCCGCCGCTGCCCTGGGGAGCTGGGACTTGGGAGCTGAGGGAGCTGAGGGCTT 240
Db 546 GCGCCGCCGCTGCCCTGGGGAGCTGGGACTTGGGAGCTGAGGGAGCTGAGGGCTT 605
Qy 241 CCACCTGGCCATCGAGCTGGA 261
Db 606 CCACCTGGCCATCGAGCTGGA 626

```

## RESULT 9

```

US-10-112-944-63
; Sequence 63, Application US/10112944
; Publication No. US20040048249A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Weng, Gezhi
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Peiyun
; APPLICANT: Xue, Aiddong J.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghost, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Zhiwei
TITLE OF INVENTION: No. US20040048249A1 Nucleic Acids and
FILE REFERENCE: 805A
CURRENT APPLICATION NUMBER: US/10/112,944
CURRENT FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25/491,404
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03/496,914
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28/515,126
PRIOR APPLICATION NUMBER: US 09/519,705
PRIOR FILING DATE: 2000-03-07/519,705
PRIOR APPLICATION NUMBER: US 09/540,217
PRIOR FILING DATE: 2000-03-31/540,217

```

## RESULT 9

```

US-10-224-624-7
; Sequence 7, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glylastoma Multiforme Associated Protein GLTEN
; FILE REFERENCE: 54704-8059 US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 7
; OTHER INFORMATION:
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(3639)
; OTHER INFORMATION:
; US-10-224-624-7

```

```

Query Match 100.0%; Score: 261; DB 15; Length: 3832;
Best Local Similarity 100.0%; Pred. No. 4.4e-123; Mismatches 0; Indels 0; Gaps 0;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAGTTGAGAATGGAGCTGAGACCAAGAGCCGGCGCCGTGCTGAGGG 60
Db 366 GATCAGTTGAGAATGGAGCTGAGACCAAGAGCCGGCGCCGTGCTGAGGG 425
Qy 61 GCTGAGGCCTGGGGAGCTGGTTCCTGGGAGAACGCTGGCCCTCATCAAGACCATAAGGGCT 120
Db 426 GCTGAGGCCTGGGGAGCTGGTTCCTGGGAGAACGCTGGCCCTCATCAAGACCATAAGGGCT 485
Qy 121 CGTGCGGCCAGCTCGCTGGGCCACCGGGAGGCCGATGCCCTGAAAGGGAGSGCCT 180
Db 486 CGTGCGGCCAGCTCGCTGGGCCACCGGGAGGCCGATGCCCTGAAAGGGAGSGCCT 545
Qy 181 GCGCCGCCGCTGCCCTGGGGAGCTGGGACTTGGGAGCTGAGGGAGCTGAGGGCTT 240
Db 546 GCGCCGCCGCTGCCCTGGGGAGCTGGGACTTGGGAGCTGAGGGAGCTGAGGGCTT 605
Qy 241 CCACCTGGCCATCGAGCTGGA 261
Db 606 CCACCTGGCCATCGAGCTGGA 626

```

FILE REFERENCE: 21402-533A

; CURRENT APPLICATION NUMBER: US/10/336,603A  
; CURRENT FILING DATE: 2003-01-03  
; PRIOR APPLICATION NUMBER: 09/746,491  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 10/055,569  
; PRIOR FILING DATE: 2001-10-26  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: CuraseqList version 0.1  
; SEQ ID NO: 25  
; LENGTH: 4801  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE: CDS  
; NAME/KEY: CDS  
; LOCATION: (178) .. (3639)  
; US-10-336-603A-25

Query Match 100.0%; Score 261; DB 12; Length 4801;  
Best Local Similarity 100.0%; Pred. No. 4.2e-123;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGACTTCGAGGAGCTGCTGAGCCAAAGACCATCGGGCT 60  
Db 543 GATCAAGGTGGACTTCGAGGAGCTGCTGAGCCAAAGACCATCGGGCT 602

Qy 61 GCTGAGCTGGGGGACCTCTGGGGAGGGGCGCTCATCAAGACCATCGGGCT 120  
Db 603 GCTGAGCTGGGGGACCTCTGGGGAGGGGCGCTCATCAAGACCATCGGGCT 662

Qy 121 CGTGCGGCGGACTGCGGCCACCCGGGAGGCCGATGCCCTGAAGGGAGGGCT 180  
Db 663 CGTGCGGCGGACTGCGGCCACCCGGGAGGCCGATGCCCTGAAGGGAGGGCT 722

Qy 181 GCCCCCGCCGCTGGCCACCGGGAGGCCGATGCCCTGAAGGGAGGGCT 240  
Db 723 GCCCCCGCCGCTGGCCACCGGGAGGCCGATGCCCTGAAGGGAGGGCT 782

Qy 241 CACCTGGCCATCGACGTGGA 261  
Db 783 CACCTGGCCATCGACGTGGA 803

RESULT 11  
US-10-276-774-950  
Sequence 950, Application US/10276774  
Publication No. US20040053245A1  
GENERAL INFORMATION:  
APPLICANT: Hysseq, Inc.  
TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides  
FILE REFERENCE: 21272-2,030  
CURRENT APPLICATION NUMBER: US/10/276,774  
CURRENT FILING DATE: 2002-11-18  
PRIOR APPLICATION NUMBER: 09/560,875  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 09/496,914  
PRIOR FILING DATE: 2000-02-03  
SOFTWARE: Custom  
SEQ ID NO: 950  
LENGTH: 4470  
TYPE: DNA  
ORGANISM: Homo sapiens

Query Match 100.0%; Score 261; DB 13; Length 4470;  
Best Local Similarity 100.0%; Pred. No. 4.3e-123;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGGTGGAGTTGAGGAGCTGCTGAGACCCAGACGGCCGGCTCTGGAGGG 60  
Db 366 GATCAAGGTGGAGTTGAGGAGCTGCTGAGACCCAGACGGCCGGCTCTGGAGGG 425

Qy 61 GCTGAGCTGGGGAGCTGTTCTGGGGAGACGGCTCCCTCATCAAGACCATCGGGCT 120  
Db 426 GCTGAGCTGGGGAGCTGTTCTGGGGAGACGGCTCCCTCATCAAGACCATCGGGCT 485

Qy 121 CGTGCGGCGGACTGCGGCCACCCGGGAGGCCGATGCCCTGAAGGGAGGGCT 180  
Db 486 CGTGCGGCGGACTGCGGCCACCCGGGAGGCCGATGCCCTGAAGGGAGGGCT 545

Qy 181 GCCCCCGCCGCTGGCCACCGGGAGGCCGATGCCCTGAAGGGAGGGCT 240  
Db 546 GCCCCCGCCGCTGGCCACCGGGAGGCCGATGCCCTGAAGGGAGGGCT 605

Qy 241 CACCTGGCCATCGACGTGGA 261  
Db 606 CACCTGGCCATCGACGTGGA 626

RESULT 12  
US-10-336-603A-25  
Sequence 25, Application US/10336603A  
Publication No. US20040072997A1  
GENERAL INFORMATION:  
APPLICANT: Alsobrook, et al.  
TITLE OF INVENTION: Tripeptidomimetic Peptides  
NUCLEIC ACIDS ENCODING SAME AND METHOD  
; CURRENT APPLICATION NUMBER: US/10/336,603A  
; CURRENT FILING DATE: 2003-01-03  
; PRIOR APPLICATION NUMBER: 09/746,491  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 10/055,569  
; PRIOR FILING DATE: 2001-10-26  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: CuraseqList version 0.1  
; SEQ ID NO: 25  
; LENGTH: 4801  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE: CDS  
; NAME/KEY: CDS  
; LOCATION: (211) .. (2883)  
; US-10-336-603A-25

Query Match 100.0%; Score 47; DB 12; Length 3847;  
Best Local Similarity 100.0%; Pred. No. 3.7e-14;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGAGGTGGAGTACAACGGGGCTTACACCTGGCATGACGTGGA 261  
Db 1 GCGAGGTGGAGTACAACGGGGCTTACACCTGGCATGACGTGGA 47

RESULT 14  
 US-10-051-769-4  
 i Sequence 4; Application US/10051769  
 i Publication No. US2003004481A1  
 i GENERAL INFORMATION:  
 i APPLICANT: MCKINNON, Randy D.  
 i TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION  
 i FILE REFERENCE: 268/260 (RNW-00-37)  
 i CURRENT APPLICATION NUMBER: US/10/051,769  
 i CURRENT FILING DATE: 2001-10-20  
 i PRIOR APPLICATION NUMBER: US 60/242,160  
 i PRIOR FILING DATE: 2000-10-20  
 i NUMBER OF SEQ ID NOS: 6  
 i SEQ ID NO 4  
 i SOFTWARE: PatentIn version 3.1  
 i LENGTH: 263  
 i TYPE: DNA  
 i ORGANISM: sprague Dawley rat  
 us-10-051-769-4

Query Match 16.5%; Score 43; DB 15; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-12;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GATCAAGTGGAGTTCGAGGAGCTGCCTGAGACCAAGACGGCC 43  
 Db 1 GATCAAGTGGAGTTCGAGGAGCTGCCTGAGACCAAGACGGCC 43

RESULT 15  
 US-10-224-624-4  
 i Sequence 4; Application US/10244624  
 i Publication No. US2003010891A1  
 i GENERAL INFORMATION:  
 i APPLICANT: Glioblastoma Multiforme Associated Protein Gliten  
 i TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten  
 i FILE REFERENCE: 54704.8059 US00  
 i CURRENT APPLICATION NUMBER: US/10/224,624  
 i CURRENT FILING DATE: 2002-08-20  
 i PRIOR APPLICATION NUMBER: 60/242,160  
 i PRIOR FILING DATE: 2000-10-20  
 i PRIOR APPLICATION NUMBER: 10/051,769  
 i PRIOR FILING DATE: 2001-10-20  
 i NUMBER OF SEQ ID NOS: 9  
 i SEQ ID NO 4  
 i SOFTWARE: PatentIn version 3.1  
 i LENGTH: 263  
 i TYPE: DNA  
 i ORGANISM: Homo sapiens  
 us-10-224-624-4

Query Match 16.5%; Score 43; DB 15; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-12;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GATCAAGTGGAGTTCGAGGAGCTGCCTGAGACCAAGACGGCC 43  
 Db 1 GATCAAGTGGAGTTCGAGGAGCTGCCTGAGACCAAGACGGCC 43

Search completed: July 18, 2004, 15:53:46  
 Job time : 265 secs



REMARK  
COMMENT

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgr@mail.nih.gov](mailto:cgr@mail.nih.gov)  
Tissue Procurement: DCTP/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNE)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency , Vancouver, BC, Canada  
[info@bcgsc.ca](mailto:info@bcgsc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Fan Guin,  
Leticia Hsieh, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Natasha Williams, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

**Clone distribution:** MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINN at: <http://image.llnl.gov>

**Series:** IRAL Plate: 29 Row: h Column: 17

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomeScan gene prediction.

FEATURES      - Location/Qualifiers  
source            1. .2623  
                  /organization "WISCONSIN"

Query Match      Best Local Similarity      Score 261; DB 9; Length 2623;

FEATURES	source	construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3' end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.	Location/Qualifiers	
	1..2639	/organism="Homo sapiens"		
		/mol_type="mRNA"		
		/db_xref="taxon:9606"		
		/clone="HRDPC2008816"		
		/cell_type="dermal papilla cells (HDPC)"		
		/clone_lib="HRDPC2"		
		/note="cloning vector: pME18SFL3-primary culture, dermal papilla cells"		
ORIGIN		Query Match 100.0%; Score 261; DB 9; Length 2639; Best Local Similarity 100.0%; Pred. No. 2.2e-31; Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy		1 GATCAAGGTGAGTTTCAGGAGTGTCTCAGAACAAAGGGGGGGCGCTCTGGAGGG 60		
Db		579 GATCAAGGTGAGTTTCAGGAGTGTCTCAGAACAAAGGGGGGGCGCTCTGGAGGG 638		
Qy		61 GCTGAGCTTGCGGACCGTGTCTCTGGGGAAAGGGGGGGCGCTCTGGAGGG 120		
Db		639 GCTGAGCTTGCGGACCGTGTCTCTGGGGAAAGGGGGGGCGCTCTGGAGGG 698		
Qy		121 CGTGGGCCAGGTGTGGCCTGCCAACGGGGAGCCCATGGCCCTAAAGGGGAGGGCT 180		
Db		699 CGTGGGCCAGGTGTGGCCTGCCAACGGGGAGCCCATGGCCCTAAAGGGGAGGGCT 758		
Qy		181 GCGGCCCTGCCAGGAGGTGGCTTGAGGTGAGTACAACGGGGCTT 240		
Db		759 GCGGCCCTGCCAGGAGGTGGCTTGAGGTGAGTACAACGGGGCTT 818		
Qy		241 CCACCTGCACATGACCTGG 261		
Db		819 CCACCTGCACATGACCTGGGA 839		
RESULT 4	BC028375	DEFINITION Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone BC028375)	LINEAR	PRI 07-OCT-2003
LOCUS	BC028375	3856 bp mRNA (bases 1 to 3856)		
DEFINITION	BC028375	Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone BC028375)		
ACCESSION	BC028375.1	complete cds.		
VERSION	GI:22382223			
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
	Klausner,R.D., Collins,F.S., Wagner,L., Shevchenko,A., Grouse,L.H., Derge,J.G., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schneider,C.F., Bhat,N.K., Jordan,H., Moore,T., Max,F., Wang,J., Hsieh,F., Diarchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Baha,S., Usdin,T.B., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.M., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,I.J., Hulyk,S.W., Villa-Lob,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fafeby,J., Heitton,E., Kettman,M., Madan,N.A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Butterfield,Y.S., Kitzmuller,M.I., Skalaka,U., Snailus,D.E., Schnurch,A., Schein,J.E., Jones,S.J. and Marra,M.A.			

TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
JOURNAL MEDLINE	22388257			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 3856)			
AUTHORS	Strausberg,R.			
TITLE	Submitted (23-APR-2002) National Institutes of Health, Mammalian CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)			
JOURNAL	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
NIH-MGC Project URL:	<a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>			
COMMENT	Contact: MGC help desk Email: <a href="mailto:cpabs+@mail.nih.gov">cpabs+@mail.nih.gov</a>			
COMMENT	Tissue Procurement: Miklos Palkovits, M.D., Ph.D. (NHGRI) & cdNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)			
COMMENT	DNA Sequencing by: Institute for Systems Biology <a href="http://www.systemsbio.org">http://www.systemsbio.org</a>			
COMMENT	contact: <a href="mailto:amadan@systemsbiology.org">amadan@systemsbiology.org</a>			
COMMENT	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>			
COMMENT	Series: ITRAK Plate: 34 Row: n Column: 21 This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.			
COMMENT	Location/Qualifiers			
COMMENT	1. .3856 /organism="Homo sapiens"			
COMMENT	/mol_type="mRNA"			
COMMENT	/db_xref="txon:9606"			
COMMENT	/clone="MGC:27107 IMAGE:4837939"			
COMMENT	/tissue_type="testis"			
COMMENT	/clone_id="NIH_MGC_97"			
COMMENT	/lab_host="DH10B"			
COMMENT	/note="Vector: pBluescript"			
COMMENT	1. .3856 /gene="LOC118987"			
COMMENT	/codon_start=1 /product="hypothetical protein LOC118987"			
COMMENT	/protein_id="AAH28355_1"			
COMMENT	/db_xref="GI:2238224"			
COMMENT	/translaton="MCLIM1MILASAVGSGFUTLIAQFELLYPROPERPDAARAGEGFVYKEPVGPGLRELYGGDRDEPSGAAGGGATTAATPTPRTCTYFLNATIIFLFLFBLRDALTREWTKLKVPEWTCGPEGAALPAACPEEAFAEVENYSGFHIALDVDFVGKSAYLKLIREKHTLNPYKTRKFPPYOTLQFVTPHWFPSFVEDLIDEVRSOBGRMPQLTSIVNQLKFLVKSRAVGRJRLVFRVTPHWFPSFVEDLIDEVRSOBGRMPQLTSIVNQLKHLVLTETVAPNSPAATADLQRDRLAIGGKITSLQVLKLKGQDRVLYYERPGQSNQGAVLQNFQGLENFUSSCSGCGYEEAAGTVTDSELDSEFEELDNLVSYDVRQNEFDQAOSLHSRSPRVPTLSIKPLGASPVLNKLAUGSHPLPKIOSKGKPPPLKTSITDP40VSKEPQTGSARFKPPVPPQAKVPLPSADPNQAEPDVLFVKEPKVPPLVDKSAEQAOKNDAIDDAAPQFQFLAQEVAKDVTSPKTDSSDPRQWTWESSELIYENKLGWTRASCIEDTBRACHYVINAIACRDPFEGJUICGHVSLKIEDVAIGCLATSTNTLISKURLEASPSKALVTRALNISMQKGNDKFCYGDITHFKYLKEGESDHVVTVTEKEPHLVEEVPSQGMGLTNKHSQDTONQNPMTCDYCKKKWTKAASOCMCAYCHKKCQEKCLAEFTGATDRIDTULKNLREGOETTLVILLATION,D.K., Muzny,D.M., Soedergren,E.J., Casavant,T.L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Toshiyuki,S., Carninci,P., Prange,C., Baha,S., Usdin,T.B., Toshiyuki,S., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.M., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,I.J., Hulyk,S.W., Villa-Lob,D.K., Sodergren,E.J., Lu,X., Gibbs,R.A., Fafeby,J., Heitton,E., Kettman,M., Madan,N.A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Butterfield,Y.S., Kitzmuller,M.I., Skalaka,U., Snailus,D.E., Schnurch,A., Schein,J.E., Jones,S.J. and Marra,M.A.			
COMMENT	LFGPSSES"			
COMMENT	/note="PDZ; Region: Domain present in PSD-95, Dlg, and LFGPSSES"			
misc_feature	1292..1546			

ZO-1/2. Also called DHR (DLG homologous region) or GLGF (relatively well conserved tetrapeptide in these domains). Some PDZs have been shown to bind C-terminal polypeptides." /db\_xref="ODD:smart00228"

**misc\_feature**  
/note="DAG PE-bind; Region: Phorbol esters/diacylglycerol binding domain (C1 domain). This domain is also known as the Protein kinase C conserved region 1 (C1) domain"  
/db\_xref="ODD:pfam00130"

**ORIGIN**

Query Match 100.0%; Score 261; DB 9; Length 3856;  
Best Local Similarity 100.0%; Pred. No. 2e-31;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
FEATURES source  
Qy 1 GATCAAAGTGGAGTTCAAGACCTGCTGAGACAAAGAACCCCTGGGG 60  
Db 568 GATCAAGGTGGAGTTCCAGGACTGCTCAGCCAAAGCGGGGGGG 627  
Qy 61 GCTGAGGCCTGGGACCTGTCCTGGGAGAGCTCCATCAAACCATCCGGT 120  
Db 628 GCTGAGGCCTGGGACCTGTCCTGGGAGAGCTCCATCAAACCATCCGGT 687  
Qy 121 CGTGCAGCAGCAGTGCTGGCCTGGCCACCGGGAGCCGATGCCCTGAAGGGAGGGCT 180  
Db 688 CGTGCAGCAGTGCTGGCCTGGCCACCGGGAGCCGATGCCCTGAAGGGAGGGCT 747  
Qy 181 GCGCGCCGCTGCCCCAGGAGCTGGCTTCAGGGGAGGTGAGTAACAGGGGGCTT 240  
Db 748 GCGCGCCGCTGCCCCAGGAGCTGGCTTCAGGGGAGGTGAGTAACAGGGGGCTT 807  
Qy 241 CCACCTGGCCATGCACTGG 261  
Db 808 CCACCTGGCCATGCACTGG 828

**RESULT 5**

AL359836/c AL359836 49052 bp DNA linear PRI 21-DEC-2001  
DEFINITION Human DNA sequence from clone RP11-389E6 on chromosome 10, complete sequence.

ACCESSION AL359836

VERSION 1.16

KEYWORDS RTG.

SOURCE Homo sapiens (human)

ORGANISM Mammalia; Eutheria; Primates; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 49052)

AUTHORS Smith, M.

TITLE Direct Submission

JOURNAL Submitted (21-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hungerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

COMMENT On Dec 23 2001 this sequence version replaced gi:17384082. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL: Sw:, SWISSPROT: Tr:, TREMBL: Wp:, WORMPEP: Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence

was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10>. RP11-389E6 is from the library RP11-2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>.

**FEATURES source**

Location/Qualifiers  
1. .49052  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-389E6"  
/clone\_llib="RP11-11.2"

**ORIGIN**

Query	Match	Score	261;	DB	9;	Length	49052;
	Best Local	Similarity	100.0%	Pred.	No.	1..1e-31;	
Matches	261;	Conservative	100.0%	Mismatches	0;	Gaps	0;
Qy	1	GATCAAAGTGGAGTTCAAGACCTGCTGAGACAAAGAACCCCTGGGG	60				
Db	568	GATCAAGGTGGAGTTCCAGGACTGCTCAGCCAAAGCGGGGGGG	627				
Qy	61	GCTGAGGCCTGGGACCTGTCCTGGGAGAGCTCCATCAAACCATCCGGT	120				
Db	628	GCTGAGGCCTGGGACCTGTCCTGGGAGAGCTCCATCAAACCATCCGGT	687				
Qy	121	CGTGCAGCAGCAGTGCTGGCCTGGCCACCGGGAGCCGATGCCCTGAAGGGAGGGCT	180				
Db	688	CGTGCAGCAGTGCTGGCCTGGCCACCGGGAGCCGATGCCCTGAAGGGAGGGCT	747				
Qy	181	GCGCGCCGCTGCCCCAGGAGCTGGCTTCAGGGGAGGTGAGTAACAGGGGGCTT	240				
Db	748	GCGCGCCGCTGCCCCAGGAGCTGGCTTCAGGGGAGGTGAGTAACAGGGGGCTT	807				
Qy	241	CCACCTGGCCATGCACTGG 261					
Db	808	CCACCTGGCCATGCACTGG 828					

**RESULT 6**

LOCUS	AC005887										
DEFINITION	CITB 173_i_12, complete sequence.										
VERSION	AC005887.3										
KEYWORDS	HGT.										
ORGANISM	Homo sapiens (human)										
REFERENCE	1 (bases 1 to 120578)										
AUTHORS	Smith, D.R.										
JOURNAL	Unpublished										
COMMENT		COMMENT		COMMENT		COMMENT		COMMENT			
ACCESSION	AL359836										
VERSION	1.16										
KEYWORDS	RTG.										
SOURCE	Homo sapiens										
ORGANISM	Mammalia; Eutheria; Primates; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	ORGANISM	Mammalia; Eutheria; Primates; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	ORGANISM	Mammalia; Eutheria; Primates; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	ORGANISM	Mammalia; Eutheria; Primates; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	ORGANISM	Mammalia; Eutheria; Primates; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	ORGANISM	Mammalia; Eutheria; Primates; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	SWISSPROT; Tr;, TREMBL; Wp;, WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence	REFERENCE	SWISSPROT; Tr;, TREMBL; Wp;, WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence	REFERENCE	SWISSPROT; Tr;, TREMBL; Wp;, WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence	REFERENCE	SWISSPROT; Tr;, TREMBL; Wp;, WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence	REFERENCE	SWISSPROT; Tr;, TREMBL; Wp;, WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> This sequence		
AUTHORS		AUTHORS		AUTHORS		AUTHORS		AUTHORS			
JOURNAL		JOURNAL		JOURNAL		JOURNAL		JOURNAL			
COMMENT		COMMENT		COMMENT		COMMENT		COMMENT			

Submitted (29-OCT-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02454, USA

Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02454, USA

REFERENCE	Street, Waltham, MA 02154, USA 4 (bases 1 to 120578)	Landers, T., Lehoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEvani, P., McFernan, K., Melidim, J., Meenus, L.J., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nicoll, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rose, C., Roman, J., Rosetti, M., Roy A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testayre, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
AUTHORS	Smith, D.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA	
REFERENCE	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA	
AUTHORS	Smith D.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA	
REMARK	Vector Sequence Clipped	
COMMENT	On Nov 5, 1999 this sequence version replaced gi:4314331.	
FEATURES	Location/qualifiers	
source	1..120578 "Homo sapiens" /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="10" /map="10q23" /clone_lib="CTP987SK-1173II12"	
ORIGIN		
Query Match	100.0%; Score: 261; DB: 9; Length: 120578;	
Best Local Similarity	100.0%; Pred. No.: 8.6e-12;	
Matches	261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 GATCAAAGTGGAGTTGAGGAGCTGGCTCAGAACRAGGGGGCGCTGTGGGG 60	
Db	53976 GATCAAGTGTGGAGTTGAGGAGCTGGCTCAGAACRAGGGGGCGCTGTGGGG 54035	
Qy	61 GCTGAGGCCTGCCGACGCTGTTCTGGCGGAAGGGTGCCCTCATCAAGACCATCGCGCT 120	
Db	54036 GCTGAGGCCTGCCGACGCTGTTCTGGGGAAAGGGTGCCCTCATCAAGACCATCGCGCT 54095	
Qy	121 CGTGGGCCAGATCTGTGCCCTGGCCACCGGGAGCCGATGGCCCTGAAGGGAGGGCT 180	
Db	54096 CGTGGGCCAGATCTGTGCCCTGGCCACCGGGAGCCGATGGCCCTGAAGGGAGGGCT 54155	
Qy	181 GCGGCCCCCTGCCGAGGAGCTGGCTTGGCCGAGGGTGTACAAACGGGCTT 240	
Db	54156 GCGGCCCCCTGCCGAGGAGCTGGCTTGGCCGAGGGTGTACAAACGGGCTT 54215	
Qy	241 CCACCTGCCATCGACCTGA 261	
Db	54216 CCACCTGCCATCGACCTGA 54236	
RESULT	7	
LOCUS	AC108407/c	
DEFINITION	Mus musculus clone RP24-422P10, DNA linear	
VERSION	AC108407.1	
KEYWORDS	GI:18377216	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Bukaryota; Metzoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mus musculus	
REFERENCE	1 (bases 1 to 68196)	
AUTHORS	Birren, B., Linton, L., Nunbaum, C. and Lander, E.	
TITLE	Unpublished	
JOURNAL	2 (bases 1 to 68196)	
REFERENCE	Andersson, S., Birren, B., Linton, L., Nunbaum, C., Lander, E., Ali, A., Allen, N., Butharia, E., Barna, N., Bastien, V., Boguslavskiy, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Cook, A., Cooke, P., DeAngelis, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Gallegos, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, J., Grand-Pierre, N., Hagos, B., Horton, L., Huime, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Keells, C., Lakocic, K., Lamazares, R.,	
AUTHORS	Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
COMMENT	All repeats were identified using RepeatMasker: <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a>	
JOURNAL	Whitehead Institute / MIT Center for Genome Research	
COMMENT	Center code: WIBR	
COMMENT	Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a>	
COMMENT	Contact: <a href="mailto:sequence_submissions@genome.wi.mit.edu">sequence_submissions@genome.wi.mit.edu</a>	
COMMENT	Project name: L18881	
COMMENT	Center Clone name: 422_P_10	
COMMENT	* NOTE: This record contains 85 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequencing is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.	
COMMENT	* 1 714: contig of 714 bp in length	
COMMENT	* 715 814: gap of 100 bp	
COMMENT	* 815 1520: contig of 706 bp in length	
COMMENT	* 1521 1620: gap of 100 bp	
COMMENT	* 1621 2323: contig of 703 bp in length	
COMMENT	* 2324 2423: gap of 100 bp	
COMMENT	* 2424 3134: contig of 711 bp in length	
COMMENT	* 3135 3234: gap of 100 bp	
COMMENT	* 3235 3937: contig of 703 bp in length	
COMMENT	* 3938 4037: gap of 100 bp	
COMMENT	* 4038 4739: contig of 702 bp in length	
COMMENT	* 4740 4839: gap of 100 bp	
COMMENT	* 4840 5545: contig of 724 bp in length	
COMMENT	* 5546 5645: gap of 100 bp	
COMMENT	* 5646 6352: contig of 707 bp in length	
COMMENT	* 6353 6452: gap of 100 bp	
COMMENT	* 6452 7172: contig of 720 bp in length	
COMMENT	* 7173 7272: gap of 100 bp	
COMMENT	* 7273 7996: contig of 688 bp in length	
COMMENT	* 7997 8096: gap of 100 bp	
COMMENT	* 8097 8814: contig of 718 bp in length	
COMMENT	* 8815 8916: gap of 100 bp	
COMMENT	* 8915 9629: contig of 715 bp in length	
COMMENT	* 9630 9729: gap of 100 bp	
COMMENT	* 9730 10417: contig of 688 bp in length	
COMMENT	* 10418 10517: gap of 100 bp	
COMMENT	* 10518 11220: contig of 706 bp in length	
COMMENT	* 11220 11324: gap of 100 bp	
COMMENT	* 11324 12014: contig of 691 bp in length	
COMMENT	* 12014 12111: gap of 100 bp	
COMMENT	* 12115 12816: contig of 702 bp in length	
COMMENT	* 12817 12916: gap of 100 bp	
COMMENT	* 12917 13616: contig of 700 bp in length	
COMMENT	* 13617 13716: gap of 100 bp	

13717	14422:	contig of 706 bp in length	*	43459	43558; gap of 100 bp	
14423	14522:	gap of 100 bp	*	43559	44242; contig of 684 bp in length	
14523	15234:	contig of 712 bp in length	*	44243	44342; gap of 100 bp	
15235	15334:	gap of 100 bp	*	44343	45043; contig of 701 bp in length	
15335	16043:	contig of 709 bp in length	*	45044	45143; gap of 100 bp	
16044	16143:	gap of 100 bp	*	45144	45856; contig of 713 bp in length	
16144	16859:	contig of 716 bp in length	*	45857	45956; gap of 100 bp	
16860	16959:	gap of 100 bp	*	45957	46654; contig of 688 bp in length	
16960	17669:	contig of 710 bp in length	*	46655	46754; gap of 100 bp	
17670	17769:	gap of 100 bp	*	46755	47449; contig of 695 bp in length	
17770	18460:	contig of 691 bp in length	*	47550	47549; gap of 100 bp	
18461	18560:	gap of 100 bp	*	47550	48258; contig of 709 bp in length	
18561	19269:	contig of 709 bp in length	*	48259	48358; gap of 100 bp	
19270	19369:	gap of 100 bp	*	48359	49070; contig of 712 bp in length	
19370	20070:	contig of 701 bp in length	*	49071	49170; gap of 100 bp	
20071	20170:	gap of 100 bp	*	49171	49886; contig of 716 bp in length	
20171	20874:	contig of 704 bp in length	*	49887	50675; gap of 100 bp	
20875	20974:	gap of 100 bp	*	50674	50677; contig of 687 bp in length	
20975	21684:	contig of 710 bp in length	*	50773:	50773; gap of 100 bp	
21685	21784:	gap of 100 bp	*	51472:	51472; contig of 699 bp in length	
21785	22491:	contig of 707 bp in length	*	51473	51572; gap of 100 bp	
22492	22591:	gap of 100 bp	*	51573	52288; contig of 716 bp in length	
22592	23299:	contig of 708 bp in length	*	52289	52388; gap of 100 bp	
23299	23399:	gap of 100 bp	*	52389	53075; contig of 687 bp in length	
23390	24086:	contig of 687 bp in length	*	53076	53175; gap of 100 bp	
24086	24186:	gap of 100 bp	*	53176	53879; contig of 704 bp in length	
24187	24901:	contig of 715 bp in length	*	53880	53979; gap of 100 bp	
24902	25001:	gap of 100 bp	*	53980	54675; contig of 696 bp in length	
25002	25714:	contig of 713 bp in length	*	54676	54775; gap of 100 bp	
25715	25814:	gap of 100 bp	*	54776	55477; contig of 702 bp in length	
25815	26528:	contig of 714 bp in length	*	55478	55577; gap of 100 bp	
26529	26628:	gap of 100 bp	*			
26629	27358:	contig of 730 bp in length	*			
27359	27458:	gap of 100 bp	*			
27459	28159:	contig of 701 bp in length	*			
28160	28259:	gap of 100 bp	*			
28260	28966:	contig of 707 bp in length	*			
28967	29066:	gap of 100 bp	*			
29067	29765:	contig of 699 bp in length	*			
29766	29865:	gap of 100 bp	*			
29866	30577:	contig of 712 bp in length	*			
30578	30677:	gap of 100 bp	*			
30678	31381:	contig of 704 bp in length	*			
31381	31481:	gap of 100 bp	*			
31482	32192:	contig of 711 bp in length	*			
32192	32292:	gap of 100 bp	*			
32293	32974:	contig of 682 bp in length	*			
32975	33074:	gap of 100 bp	*			
33075	33760:	contig of 686 bp in length	*			
33761	33860:	gap of 100 bp	*			
33861	34574:	contig of 714 bp in length	*			
34575	34674:	gap of 100 bp	*			
34675	35378:	contig of 704 bp in length	*			
35379	35478:	gap of 100 bp	*			
35479	36194:	contig of 716 bp in length	*			
36195	36294:	gap of 100 bp	*			
36295	36973:	contig of 679 bp in length	*			
36974	37073:	gap of 100 bp	*			
37074	37774:	contig of 701 bp in length	*			
37775	37874:	gap of 100 bp	*			
37875	38586:	contig of 712 bp in length	*			
38587	38686:	gap of 100 bp	*			
38687	39388:	contig of 702 bp in length	*			
39389	39488:	gap of 100 bp	*			
39489	40195:	contig of 707 bp in length	*			
40196	40295:	gap of 100 bp	*			
40296	41013:	contig of 718 bp in length	*			
41014	41113:	gap of 100 bp	*			
41114	41828:	contig of 715 bp in length	*			
41929	42643:	gap of 100 bp	*			
42644	42743:	contig of 715 bp in length	*			
42744	43458:	contig of 715 bp in length	*			
				RESULT 8		
				AC139040		
				LOCUS	AC139040	221524 bp
				DEFINITION	Mus musculus chromosome 19,	DNA linear
				ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	clone RP23-29ZH20, complete sequence.
				VERSION	AC139040.7	GR:32813610
				KEYWORDS	HTG,	
				SOURCE	Mus musculus (house mouse)	
				ORGANISM	Mus musculus	
				REFERENCE	1 (bases 1 to 221524)	
				AUTHORS	Birren, B., Nusbaum, C. and Lander, E.	
				TITLE	Mus musculus chromosome 19, clone RP23-29ZH20	
				JOURNAL	Unpublished	
				REFERENCE	2 (bases 1 to 221524)	



repeat_region	8336 . 8484 / rpt_family="B1_MM"	KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
repeat_region	8968 . 9042 / rpt_family="L2" complement(1939 . 9332)	SOURCE Rattus norvegicus (Norway rat)
repeat_region	/ rpt_family="MIR" complement(1953 . 9742)	ORGANISM Rattus norvegicus
repeat_region	/ rpt_family="B2_Mm2" complement(19847 . 10055) / rpt_family="B3"	MAMMALIA; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
repeat_region	11638 . 11844 / rpt_family="B3"	Rattus. 1 (bases 1 to 243412)
repeat_region	12218 . 12494 / rpt_family="RMR14" complement(13731 . 13993)	REFERENCE Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J., Allen, C., Allier, H., Alstroem, S., Amiri, A., Arguello, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswali, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buahy, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacon, J., Chavez, D., Chen, G., Chen, Y., Cree, A., D'Souza, L., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davis, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escott, M., Eugene, C., Evans, C.A., Faulls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebrégeorgis, E., Gear, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Guanatane P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandes, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howell, S., Hu, S., Huame, J., Idelbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., King, L., Kovar, C., Gunaratne P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandes, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshawa, L., Loulseged, H., Mahmud, M., Mallory, K., Mangum, A., Maheshwari, M., Mahindarase, M., Mahmud, M., Martin, R., Martinez, E., Mangum, B., Mapua, P., Martin, K., Martin, R., Martineau, S., Mawhinney, S., McNeill, T.Z., Meinen, E., McNeill, T.Z., Meinen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, D., Newton, D., Nguyen, N., Norris, S., Nwakwe, Ieme, O., Okwundu, G., Olarupunguo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, J.L., Pfannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puato, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Sawyer, G., Scheerer, S., Scott, G., Shatsman, S., Shen, H., Sheety, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Smeid, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svarek, A., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, P., Taylor, C., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Wei, X., White, F., Williams, G., Williamson, R., Wleczyk, R., Woerner, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhäusern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.
repeat_region	15345 . 15509 / rpt_family="LIME3" complement(15218 . 15218)	TITLE Direct Submission JOURNAL Unpublished AUTHORS 2 (bases 1 to 243412)
repeat_region	15850 . 15857 / rpt_family="LIMC4" complement(15704 . 15731)	REFERENCE Worley, K.C.
repeat_region	15875 . 16100 / rpt_family="ID_B1"	AUTHORS Direct Submission
repeat_region	15732 . 15849 / rpt_family="ORR1D" complement(16101 . 16251)	JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
repeat_region	17479 . 17513 / rpt_family="CA"	COMMENT On Nov 19, 2002 this sequence version replaced gi:23269487. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas
repeat_region	17626 . 17653 / rpt_family="ORR1D"	
repeat_region	18382 . 18451 / rpt_family="ID_B1" complement(18594 . 18730)	
repeat_region	240; Conservative Matches 240; Indels 0; Gaps 0; Mismatches 21;	
Query Match	87.1%; Score 227.4; DB 10; Length 221524;	
Best Local Similarity	92.0%; Pred. No. 1.2e-26;	
Matches		
Qy	1 GATCAAAGTGGGGACGTGGAGCTGGTCAGACCAAGAGCGCCGCCCTCATAGAACATCGGGCT 60	
Db	42833 GATCAAAGTGGGGACGTGGAGCTGGTCAGACCAAGAGCGCCGCCCTCATAGAACATCGGGCT 42892	
Qy	61 GCTGAGGCCTGGGGACGTGGAGCTGGTCAGACCAAGAGCGCCGCCCTCATAGAACATCGGGCT 120	
Db	42893 GCTGAGGCCTGGGGACGTGGAGCTGGTCAGACCAAGAGCGCCGCCCTCATAGAACATCGGGCT 42952	
Qy	121 CGTGGCGCCGACTGTGGCTCTGGCCACGGGAGCGCGATGCCCTGAAGGAGCGCGCT 180	
Db	42953 GGTGGCGCCGACTGTGGCTCTGGCCACGGGAGCGCGATGCCCTGAAGGAGCGCGCT 43012	
Qy	181 GCCCGCGCCGCCCCGGAGCTGGCTTCAGGGAGCTGGACTAACGGGGCT 240	
Db	43013 GCGCGGCACTGGCCGGAGCGCTGGCTTCAGGGAGCTGGACTAACGGGGCT 43072	
Qy	241 CCACCTGGCCATCGACCTGGAA 261	
Db	43073 CCACCTGGCCATCGACCTGGAA 43093	
RESULT	9	
AC127769	AC127769	
LOCUS	243412 bp	DNA linear
DEFINITION	Rattus norvegicus clone CH23-0-15/C16,	WORKING DRAFT SEQUENCE, 3
UNORDERED	unordered pieces.	
ACCESSION	AC127769	
VERSION	AC127769 . 3	GI: 25079525

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GZSV

Center clone name: CH211-157C16

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 220496 bases at least Q40

Consensus quality: 222898 bases at least Q30

Consensus quality: 224623 bases at least Q20

Estimated insert size: 224491; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length.  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank/draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 11919: contig of 11919 bp in length  
\* 12019: gap of unknown length  
\* 12020 18301: contig of 6282 bp in length  
\* 18302 18401: gap of unknown length  
\* 18402 243412: contig of 225011 bp in length.  
Location/Qualifiers  
1..243412

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="Taxon:10116"

/clone="CH211-157C16"

1..1535

/note="wgs\_contig"

12020..13456

/note="wgs\_contig"

17209..18301

/note="wgs\_contig"

18402..19412

/note="wgs\_contig"

25075..27269

/note="wgs\_contig"

238399..242288

/note="wgs\_contig"

242339..243412

/note="wgs\_contig"

FEATURES

source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

ORIGIN

Query Match

Best Local Similarity

Matches

2337; conservative

0;

Mismatches

0;

Indels

0;

Gaps

0;

FEATURES

source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

ORIGIN

Query Match

Best Local Similarity

Matches

2337; conservative

0;

Mismatches

0;

Indels

0;

Gaps

0;

Db	90890	GCTGAGCCCTGGCGAACGGAGCTGTGCAAGACCAAGAAGCCGCCTCATACAGACCATCCGGCT	90949
Qy	121	CGTGCGCCAGTCGTGCGCTTGGCCACGCCGAGCCGATGCCCTGAAGGGAGGGCT	180
Db	90950	GGTGCSCCCCTGTTGCGTCGACGGCGACCGAACCCGACTCT	91009
Qy	181	GCCGCGCCCTGCCGGAGCTGCCTTGAGGGAGGTGAGTACAAGGGGGCT	240
Db	91010	GCGCGACCTGCCGGAGGGCTGCCTTGAGGGAGGTGAGTACAAGGGGGCT	91069
Qy	241	CCACCTGGCCATCGACTGGA	261
Db	91070	CCACCTGGCCATCGACTGGA	91090
RESULT 10			
LOCUS	BX511030	179734 bp	DNA linear
DEFINITION	Danio rerio clone CH211-6612B	WORKING DRAFT SEQUENCE	3 unordered pieces.
ACCESSION	BX511030	GI:323919530	
VERSION	HTGS DRAFT	HTGS_FULLTOP	
KEYWORDS	HTGS PHASE1	HTGS	
SOURCE	Danio rerio	(zebrafish)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Buteleostomi; Cypriniformes; Cyprinidae; Danio; McLaren,S.		
REFERENCE	1 (bases 1 to 179734)		
AUTHORS	McLaren,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk		
COMMENT	On Jul 1, 2003 this sequence version replaced gi:3107341.		
Center: Wellcome Trust Sanger Institute			
Center code: SC			
Web site: <a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a>			
Contact: zfish-help@sanger.ac.uk			
----- Project Information			
Center project name: GZSV			
Center clone name: CH211-157C16			
----- Summary Statistics			
Assembly program: Phrap; version 0.990329			
Consensus quality: 220496 bases at least Q40			
Consensus quality: 222898 bases at least Q30			
Consensus quality: 224623 bases at least Q20			
Estimated insert size: 224491; sum-of-contigs estimation			
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation			
* NOTE: Estimated insert size may differ from sequence length. * (see <a href="http://www.hgsc.bcm.tmc.edu/docs/Genbank/draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/Genbank/draft_data.html</a> ).			
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.			
* 1 11919: contig of 11919 bp in length			
* 12019: gap of unknown length			
* 12020 18301: contig of 6282 bp in length			
* 18302 18401: gap of unknown length			
* 18402 243412: contig of 225011 bp in length.			
Location/Qualifiers			
1..243412			
/organism="Rattus norvegicus"			
/mol_type="genomic DNA"			
/db_xref="Taxon:10116"			
/clone="CH211-157C16"			
1..1535			
/note="wgs_contig"			
12020..13456			
/note="wgs_contig"			
17209..18301			
/note="wgs_contig"			
18402..19412			
/note="wgs_contig"			
25075..27269			
/note="wgs_contig"			
238399..242288			
/note="wgs_contig"			
242339..243412			
/note="wgs_contig"			
FEATURES			
source			
misc_feature			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
2337; conservative			
0;			
Mismatches			
0;			
Indels			
0;			
Gaps			
0;			
FEATURES			
source			
misc_feature			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
2337; conservative			
0;			
Mismatches			
0;			
Indels			
0;			
Gaps			
0;			
FEATURES			
source			
misc_feature			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
2337; conservative			
0;			
Mismatches			
0;			
Indels			
0;			
Gaps			
0;			
FEATURES			
source			
misc_feature			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
90830			
GATCAAGGCTGGCTGGGAGCTGTCGACAGCTGAGCTGCTGCGCTTCAAGACCACGGGT			
90889			
GCTGAGCCCTGGCTGGGAGCTGTCGACAGCTGAGCTGCTGCGCTTCAAGACCACGGGT			
91			

```

misc_feature
  1. . 89039
    /note="assembly fragment :02405
      fragment chain:1
      clone end:T7
      vector side:left"
    89140_.168174

  misc_feature
    /note="assembly fragment :01570
      fragment chain:1
      clone end:SP6
      vector side:right"
    169275_.179734

  misc_feature
    /note="assembly fragment :00769"
      ORIGIN
        Query Match 33.6% Score 87.6; DB 2; Length 179734;
        Best Local Similarity 78.4%; Pred. No. 5..1e-05;
        Matches 105; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
        /db 1 GATCAAGGTGGAGTTCAAGGAGTGTCTAGACCAAGCAGCCGGCTGCTGGAGGG 60
        QY 167240 GATCAAGGTAGAGTTCAAGGACTCTCAGACCAAGCTGCAGGGCTGCTGGAGGG 167299
        QY 61 GCTGAGGCTGCCGAAGCTTGCTGGCGAGACGGTGCCTTCATCAAACCATCCGGCT 120
        Dbd 167300 GCTGAGTGTGGGAGTGTCCCTGGAAACTGGTGCCTGGTCAAAACCCGAGCT 167359
        QY 121 CGTGGGCCAGTCG 134
        Dbd 167360 AATGAAACCCGTGG 167373

      RESULT 11
      AFA469049/c
      LOCUS AFA469049 18666 bp DNA linear VRT 18-APR-2002
      DEFINITION Gallus gallus caspase-6 gene, complete cds.
      ACCESSION AFA469049
      VERSION AFA469049.1 GR:18920651
      SOURCE Gallus gallus (chicken)
      ORGANISM Gallus gallus
      KEYWORDS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Phasianinae; Aves; Neognathae; Galliformes; phasianidae;
      AUTHORS Ruchaud,S.,Korfali,N.,Villa,P.,Dingwall,C.,Kaufmann,S.H.,and Earnshaw,W.C.
      TITLE Caspase-6 gene disruption reveals a requirement for lamin A cleavage in apoptotic chromatin condensation
      JOURNAL EMBO J. 21 (8), 1967-1977 (2002)
      MEDLINE 21950342
      PUBMED 11953316
      REFERENCE 2 (bases 1 to 18666)
      AUTHORS Ruchaud,S.,Korfali,N.,Villa,P. and Earnshaw,W.C.
      TITLE Direct Submission
      JOURNAL Submitted (16-JAN-2002) ICMB, Edinburgh University, Mayfield Road,
      Edinburgh EH9 3JR, UK
      FEATURES source
      /organism="Gallus gallus"
      /db_xref="taxon:9031"
      /product="caspase-6"
      /codon_start=1
      /product="caspase-6"
      /protein_id="P18920652"
      /db_xref="GI:18920652"
      /translation="MSGAERRRAAAGRVOILSDKSPTPTTAADGNQNITEVDAAFDKRTTFD
      PFEQKXNQQRRIIIFNDDTSNAQDFCVCLSLDGTAYDAQOLKLTIDTNNMFGRFIDKK
      LRAEDVLLKKVFSRQDTSNAQDFCVCLSLDGTAYDAQOLKLTIDTNNMFGRFIDKK
      FDK
```

surrounding genes and the short region (67-162aa) of similarity to TR:052357 (EMBL:J38472) terD, associated with phage inhibition, colicin resistance and tellurite resistance from *Serratia marcescens* IncH12 plasmid R478. BLAST scores = E=1e-12, Identities = 40/98 (40%), positives = 55/98 (56%) which may represent an active

```
/codon_start=1  
/trans_table=11  
/product="ABC excision nuclease subunit B"  
/protein_id="CD55287_1"  
/db_xref="IGI:24419016"  
/db_xref="QGOA:QBCK11"
```

/db\_xref="SPREMBI:Q93IJ3"  
 /translation="MHYRAAVTTALLGVALTAPLSHARADEAGDDGPTVYAHGAS  
 GYAPENTAAVDPAAELGTVWENDVORTDGLUVVHDESLARTTUVBETPDRSPW  
 KVDFDTAEALRIDLADSGMSEYAGARTPLEYVHRVEFHQKLLELRSGLYPGI  
 EOTLKVIANEWGLDRREHAGRLVQSFDAIDSRTVHDLPKAVTKGFLGTAVSELP  
 Query Match 20.8%; Score 54.4%; DB 1; Length 321250;  
 Best Local Similarity 52.7%; Pred. No. 6;  
 Matches 118; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
 Title JOURNAL  
 Reference AUTHORS  
 QY 4 CAAGGTGGAGTTGCGGACGTCAGACCAAGGGGGCTGTGGAGGGGCT 63  
 DB 204911 CGAGGGCGGCACGAGCTGCCGGGGTGAACGGCTGCACGGATCATCGGGCCGAT 204970  
 Title JOURNAL  
 Submitted (25-SEP-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 Direct Submission  
 Worley, K.C.  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: TUGU  
 Center clone name: CH230-1K3  
 Summary Statistics  
 Sequencing Vector: Plasmid  
 Chemistry: Dye-Terminator Big Dye: 100% of reads  
 Assembly program: phrap; version 0.990329  
 Consensus quality: 28004 bases at least Q30  
 Consensus quality: 31728 bases at least Q30  
 Consensus quality: 33612 bases at least Q20  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see <http://www.hgsc.bcm.tmc.edu/docs/Genbank/draft.html>).  
 \* NOTE: This is a working draft sequence. It currently  
 \* consists of 39 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

RESULT 13  
 AC134262 Locus 67702 bp DNA linear HTG 25-SEP-2002  
 DEFINITION Rattus norvegicus clone CH230-1K3, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 39 unordered pieces.  
 AC134262 Version GI:23307916  
 KEYWORDS HTG, HTGS,\_PHASE1.  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Butcheria; Rodentia; Sciurognathii; Muridae; Murinae;  
 Rattus.  
 1. (bases 1 to 67702)  
 MUZNY,D.M., ADAMS,C., ADIO-OUDOLA,B., ALI-OSSMAN,F.R., ALLEN,C.,  
 ALSBROOKS,S.L., AMARATUNGE,H.C., ARE,J.R., AYELE,M., BANKS,T.,  
 BARBARIA,J., BENTON,J., BIMAGE,K., BLANKENSHIP,K., BONNIN,D.,  
 BOUCK,J., BOWIE,S., BRIEVA,M., BROWN,E., BROWN,M., BRYANT,N.P.,  
 BURAY,C., BURCH,P., BURKEIT,K., BURRELL,C., BYRD,N.C.,  
 CARRON,T.F., CARTER,M., CAVAZOS,S.R., CHACKO,J.J., CHAVEZ,D.,  
 CHEN,G., CHEN,R., CHEN,Z., CHOWDHURY,I., CHRISTOPOLOUS,C.,  
 CLEVELAND,C.D., COX,C., COYLE,M.D., DAETHORNE,S.R., DAVID,R.,  
 DAVID,L., DAVIS,C., DARY-CARROLL,L., DEDERICH,D.A.,  
 DELANEY,K.R., DELGADO,O., DENIA,A.L., DIRIG,Y., DINI,H.H.,  
 DOWTHWAITE,K.J., DRAPER,H., DUGAN-ROCHA,S., DURBIR,K.J.,  
 EARNHARDT,C., EDGAR,D., EDWARDS,C.C., ELBAJ,C., ESCOTT,M.,  
 FELLIS,T., FERRAGUTO,D., FLAIG,N., FORD,J., FOSTER,P., FRANTZ,P.,  
 GABISA,A., GAO,J., GARCIA,A., GARNER,T., GARZA,N., GILL,R.,  
 GORELL,J.H., GUEVARA,W., GUNARATNE,P., HALE,S., HAMILTON,K.,  
 HARRIS,C., HARRIS,K., HART,M., HAVILAK,P., HAWES,A., HERNANDEZ,J.,  
 HOMSI,F., HOWARD,S., HUBER,J., HULPK,S., HUME,J., HOLLINS,B.,  
 JACOBSON,B., JIA,Y., JOHNSON,R., JOLIVET,S., JOUTAH,S.,  
 KRATZOVIC,J., KURESHI,A., LANDRY,N., LEAL,B., LEWIS,L.C., LEWIS,L.,  
 LI,J., LI,Z., LIGHTERGE,O., LIEU,C., LIU,J., LIU,W., LOULSEGHD,H.,  
 LOZADO,R.J., LU,X., LUCIER,A., LUCIER,R., LUNA,R., MA,J.,  
 MAHESHWARI,M., MAPUA,P., MARTIN,R., MARTINDALE,A., MARTINEZ,E.,  
 MASSEY,E., MAWHINNEY,E., MCLEOD,M.P., MEDDOR,M., MEI,G., METZKER,M.,  
 MINER,Z., MITCHELL,T., MOHABBAT,K., MORGAN,M., MORRIS,S.,  
 MOSER,M., NEAL,D., NEWTON,J., NEWTON,N., NGUYEN,A., NGUYEN,N.,  
 NGUYEN,N., NICKERSON,E., NWOKONKWO,S., OGUCHI,M., OKUONU,G.,  
 ORAGUNYE,N., OVIEDO,R., PACE,A., PAYTON,B., PEERY,J., PEREZ,L.,  
 PETERS,L., PICKENS,R., PRIMUS,E., PU,L.I., QUILES,M., REN,Y.,  
 RIVES,M., ROJAS,A., ROJABOKAN,I., ROLFE,M., RUIZ,S., SAVERY,G.,  
 SCHEFER,S., SCOTT,G., SHEN,H., SISON,I., SISSON,I.



REFERENCE 3 (bases 1 to 219992) Best Local Similarity 52.0%; Pred. No. 17;  
 AUTHORS Rat Genome Sequencing Consortium. Matches 116; Conservative 0; Mismatches 107; Indels 0; Gaps 0;  
 TITLE Direct Submission  
 JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On Nov 6, 2002 this sequence version replaced gi:23682583.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole genome  
 shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GTCW  
 Center clone name: CH230-142114  
 ----- Summary Statistics  
 Assembly program: phrap; version 0.990329  
 Consensus quality: 20/077 bases at least Q40  
 Consensus Quality: 209387 bases at least Q30  
 Consensus Quality: 210977 bases at least Q20  
 Estimated insert size: 21380; sum-of-contigs estimation  
 Quality coverage: 9x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 218703: contig of 218703 bp in length  
 \* 218803: gap of unknown length  
 \* 219992: contig of 1189 bp in length.  
 FEATURES Location/Qualifiers  
 source 1..219992  
 /organism=Rattus norvegicus  
 /mol\_type="genomic DNA"  
 /db\_xref="eSaxon-10116"  
 /clone="CH230-142114"  
 .1..1360  
 /note="wgs end\_extension  
 clone\_end:5p6"  
 misc\_feature 2228..3263  
 /note="wgs end\_extension  
 clone\_end:5p6"  
 misc\_feature 10120..10615  
 /note="wgs end\_boundary  
 clone\_end:5p6"  
 misc\_feature 201759..217135  
 /note="clone\_boundary  
 clone\_end:r7"  
 misc\_feature 19..88; Score 51.8; DB 2; Length 219992;  
 ORIGIN end\_sequence:RWAXN55Tvb"  
 Query Match Shetty,J., Shartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Rivers,C., Rodkey,T., Rojas,A., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scheirer,S., Scott,G., Shatsman,S., Shen,H.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P., Rives,C., Rodkey,T., Scott,G., Shartsman,S., Shen,H.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Narkiewicz,C., Neal,D., Newton,N., Norris,S., Nwaokelemech,O., Okonkwo,G., Olarunpungsagon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,L., Pfnannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,B., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Mawhinney,S., McLeod,M.P., McNeill,T.Z., Montemayor,J., Moore,S., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lohesuweva,L., Louie,H., Lozado,R.J., Lu,X., Ma,J., Mahindartne,M., Mahmud,M., Mallory,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Kowis,C., Kraft,C.L., Lebow,H., Khan,Z., Khan,Z., Kovari,C., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlik,P., Haves,A., Hernandez,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howell,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Leiby,S., Khan,Z., King,L., Kovari,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lohesuweva,L., Louie,H., Lozado,R.J., Lu,X., Ma,J., Mahindartne,M., Mahmud,M., Mallory,K., Mangum,A., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Narkiewicz,C., Neal,D., Newton,N., Norris,S., Nwaokelemech,O., Okonkwo,G., Olarunpungsagon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,L., Pfnannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,B., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P., Rives,C., Rodkey,T., Rojas,A., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scheirer,S., Scott,G., Shartsman,S., Shen,H., Shetty,J., Shartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,



